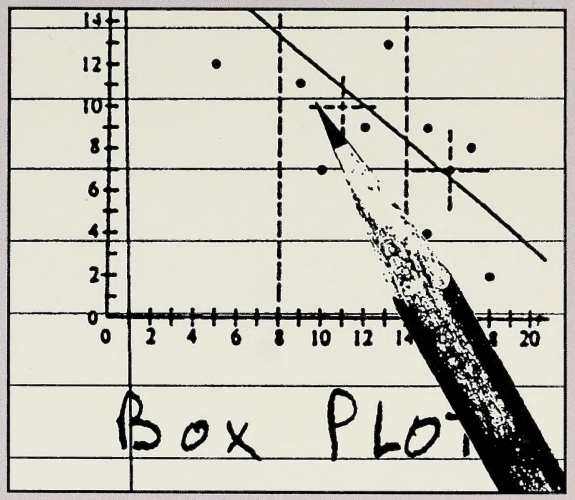
MATHEMATICS 33

Statistics



Unit 6

Distance Learning

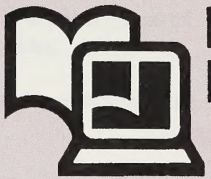
Alberta EDUCATION

University of Alberta Library
0 1620 3409880 4

Digitized by the Internet Archive
in 2016 with funding from
University of Alberta Libraries

https://archive.org/details/mathematics3306albe

Welcome



Distance Learning

You have chosen an alternate form of learning that allows you to work at your own pace. You will be responsible for your own schedule, for disciplining yourself to study the units thoroughly, and for completing your units regularly. We wish you much success and enjoyment in your studies.

Mathematics 33 Student Module Unit 6 Statistics Alberta Distance Learning Centre ISBN No. 0-7741-0170-9

Copyright © 1991, the crown in Right of Alberta, as represented by the Minister of Education, 11160 Jasper Avenue, Edmonton, Alberta, T5L 0L2. Further reproduction of this material by any duplicating process without written permission of Alberta Education is a violation of copyright.

No part of this courseware may be reproduced in any form, including photocopying (unless otherwise indicated), without the written permission of Alberta Education. Additional copies may be obtained from the Learning Resources Distributing Centre, 12360 - 142 Street, Edmonton, Alberta, T5L 4X9.

Care has been taken to trace ownership of copyright material. Any information which will enable Alberta Education to rectify any reference or credit in subsequent printings will be gladly received by the Director, Alberta Distance Learning Centre, Box 4000, Barrhead, Alberta, T0G 2P0.

UNIVERSITY LIBRARY
UNIVERSITY OF ALBERTA

General Information

This information explains the basic layout of each booklet.

- **What You Already Know** and **Review** are to help you look back at what you have previously studied. The questions are to jog your memory and to prepare you for the learning that is going to happen in this unit.
- As you begin each **Topic**, spend a little time looking over the components. Doing this will give you a preview of what will be covered in the topic and will set your mind in the direction of learning.
- **Exploring the Topic** includes the objectives, concept development, and activities for each objective. Use your own papers to arrive at the answers in the activities.
- **Extra Help** reviews the topic. If you had any difficulty with **Exploring the Topic**, you may find this part helpful.
- **Extensions** gives you the opportunity to take the topic one step further.
- To summarize what you have learned, and to find instructions on doing the unit assignment, turn to the **Unit Summary** at the end of the unit.
- The **Appendices** include the solutions to **Activities (Appendix A)** and any other charts, tables, etc. which may be referred to in the topics **(Appendix B, etc.)**.

Visual Cues

Visual cues are pictures that are used to identify important areas of the material. They are found throughout the booklet.

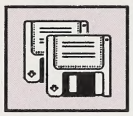
An explanation of what they mean is written beside each visual cue.

Key Idea
- flagging important ideas

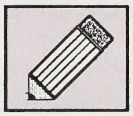
Audiotape
- learning by listening to an audiotape

What You Already Know
- reviewing what you already know

Another View
- exploring different perspectives



Computer Software
- learning by using computer software



Review
- studying previous concepts

Solutions
- correcting the activities

Videotape
- learning by viewing a videotape

Introduction
- introducing the unit

Extra Help
- providing additional study

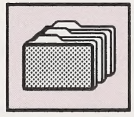
Print Pathway
- choosing a print alternative

What Lies Ahead
- previewing the unit

Extensions
- going on with the topic

Calculator
- using your calculator

Exploring the Topic
- actively learning new concepts



What You Have Learned
- summarizing what you have learned

Mathematics 33

Course Overview

Mathematics 33 contains 8 units. Beside each unit is a percentage that indicates what the unit is worth in relation to the rest of the course. The units and their percentages are listed below. You will be studying the unit that is shaded.

Unit 1 Powers and Radicals	10%
Unit 2 Polynomials and Rational Expressions	10%
Unit 3 Functions and Relations	16%
Unit 4 Quadratic Functions and Equations	20%
Unit 5 Trigonometry	16%
Unit 6 Statistics	**16%**
Unit 7 Annuities	6%
Unit 8 Mortgages and Loans	6%
	100%

Unit Assessment

After completing the unit you will be given a mark based totally on a unit assignment. This assignment will be found in the Assignment Booklet.

Unit Assignment - 100%

If you are working on a CML terminal your teacher will determine what this assessment will be. It may be

Unit Assignment - 50%
Supervised Unit Test - 50%

Introduction to Statistics

This unit covers topics dealing with statistics. Each topic contains explanations, examples, and activities to assist you in understanding statistics. If you find you are having difficulty with the explanations and the way the material is presented, there is a section called **Extra Help**. If you would like to extend your knowledge of the topic, there is a section called **Extensions**.

You can evaluate your understanding of each topic by working through the activities. Answers are found in the solutions in **Appendix A**. In several cases there is more than one way to do the question.

Unit 6 Statistics

Contents at a Glance

Value	**Statistics**	3
	What You Already Know	5
	Review	5
50%	**Topic 1:** Bivariate Data	7
	• Introduction • Extra Help • What Lies Ahead • Extensions • Exploring Topic 1	
50%	**Topic 2:** Confidence	45
	• Introduction • Extra Help • What Lies Ahead • Extensions • Exploring Topic 2	
	Unit Summary	89
	• What You Have Learned • Unit Assignment	
	Appendices	90
	• Appendix A • Appendix B	

Statistics

You are about to discover the word statistics. Try opening a newspaper without seeing statistics. It is almost impossible. Statistics are everywhere, and they are used by everyone for many reasons. Statisticians often try to determine whether there are relationships between variables. For example, if you eat breakfast, are you likely to live longer? If you study for two hours a day, will your math marks increase?

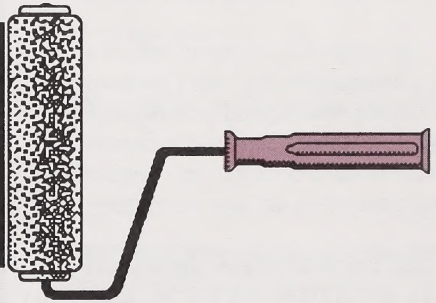
Statisticians also try to estimate the outcome of important events. For example, surveys are often used to see who is favoured in an election, or to see if the public believes attendance at professional hockey games will decrease without the team's star player. Pollsters reflect public opinion based on a sample of the entire population.

You may not become a statistician by doing this unit; however, your ability to interpret the statistics you encounter in everyday life will greatly improve.

Unit 6 Statistics

Topic 1 Bivariate Data

Topic 2 Confidence



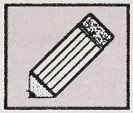
What You Already Know

Refresh your memory!

You should be able to do the following:

- Given a set of data, determine the mean.
- Given a set of data, determine the median.
- Given two points on a line, determine the slope of the line.
- Given two points on a line, determine the equation of the line.
- Given an ordered pair(s), plot the ordered pair(s) on a Cartesian plane.
- Given a line, estimate the ordered pairs of two points on the line.
- Given a ratio, determine the equivalent percentage.

Now that you have looked at material that you studied previously, go to the **Review** to confirm your understanding of this material.



Review

Try the following review questions.

1. Find the mean for the following groups of data.
   a. 18, 34, 14
   b. 6, 4, 7, 10, 9, 9, 11, 16
   c. 141, 107, 118, 114, 125
2. Find the median for the following groups of data.
   a. 18, 34, 14
   b. 6, 4, 7, 10, 9, 9, 11, 16
   c. 141, 107, 118, 114, 125
3. Determine the slope of a line that passes through the following pairs of points.
   a. $(4, 7)$ and $(9, 5)$
   b. $(108, 22)$ and $(122, 102)$

4. Determine the equation of the line that passes through the following pairs of points.

   a. $(4, 7)$ and $(9, 5)$

   b. $(108, 22)$ and $(122, 102)$

5. Plot the following ordered pairs on a Cartesian plane.

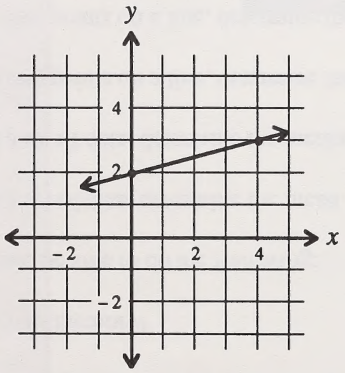
   a. $A(3, 4)$

   b. $B(-2, 5)$

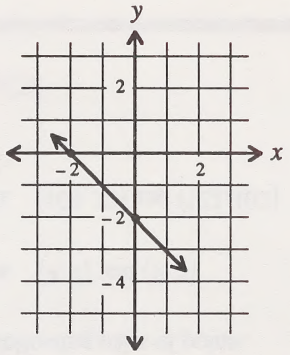
   c. $C(-5, 0)$

6. Find the slope of the following lines.

   a.



   b.



7. Convert the following ratios to percent.

   a. $\frac{12}{25}$

   b. $\frac{17}{20}$

   c. $\frac{9}{60}$

Now go to the **Review** solutions in **Appendix A.**

If you feel comfortable with this material, then you are ready to move on to the first topic. If you had trouble with this material, then review the Statistics unit in **Mathematics 13** and the Linear Relations unit of **Mathematics 23** before proceeding to the first topic.

Topic 1 Bivariate Data

Introduction

Usually you are interested in how a certain characteristic influences the outcome of a situation.

How does the speed you drive influence your gas mileage?

How does the diameter of a cable influence the amount of mass it can support?

How does the speed of a running back in football influence the number of yards rushed in a season?

Each of the above situations is made up of data that has two numbers. In the first case, for every speed that you test, you will get a specified gas mileage. These two numbers (speed and gas mileage) can be used to plot a single point on a graph. If enough points are plotted on a graph, a pattern may become evident. The pattern is what you are going to be looking for in this topic.

What Lies Ahead

Throughout this topic you will learn to

1. plot sets of bivariate data on a scatterplot
2. plot a line of best fit on a scatterplot using the median fit method
3. use the equation of the line of best fit to generate new data for a population
4. determine the strength and type of correlation between the variables of a bivariate distribution
5. collect, organize, and analyze sets of bivariate data

Now that you know what to expect, turn the page to begin your study of bivariate data.

Exploring Topic 1

Activity 1

Plot sets of bivariate data on a scatterplot.

In studying statistics to this point, you have studied data that involved only one variable. However, it is frequently more valuable to study data that involves two variables. Data that consists of two variables is called bivariate data.

When bivariate data is plotted on a scatterplot, patterns may become evident that are not visible when examining the data alone.

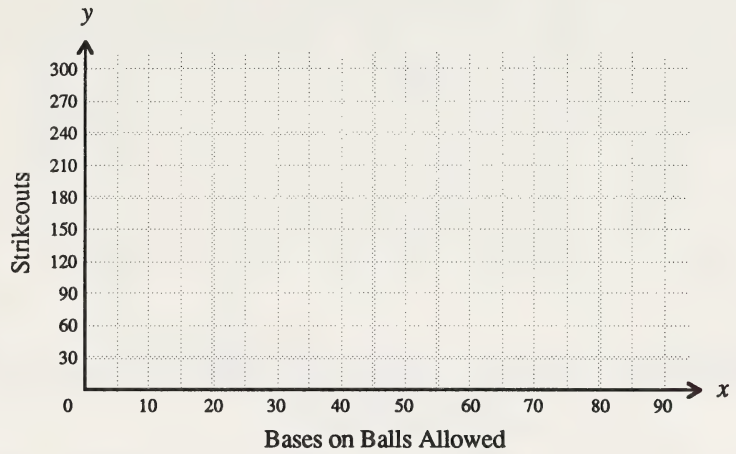
Use an example to plot and analyze a scatterplot.

The following chart contains the number of strikeouts and the number of bases on balls allowed by a number of pitchers for the 1989 season in the Distance Baseball League.

Pitcher	Bases on Balls	Strikeouts
Maddon, Calgary	81	140
Prune, Edmonton	44	131
Conic, Lethbridge	80	213
Daze, Medicine Hat	86	298
Garney, Grande Prairie	89	162
Drysday, High Level	50	127
Shoulder, Lloydminster	53	144
Jakson, Drumheller	71	161
Reussel, Red Deer	42	92
Hornblower, Jasper	73	178
Scottie, Banff	54	190
Suttuer, Calgary	70	144
Martten, Edmonton	55	120
Coodie, Lethbridge	57	175
Terence, Medicine Hat	34	65
Carrie, Grande Prairie	68	116
Smilie, Stettler	46	129
Mussen, St. Paul	58	112
Browner, St. Albert	64	124
Drowns, Wetaskiwin	47	118
Larry, Banff	56	180
Knocks, Red Deer	67	103

Step 1: Decide on which axis each variable will be placed. Decide on the scale for each axis and label each axis.

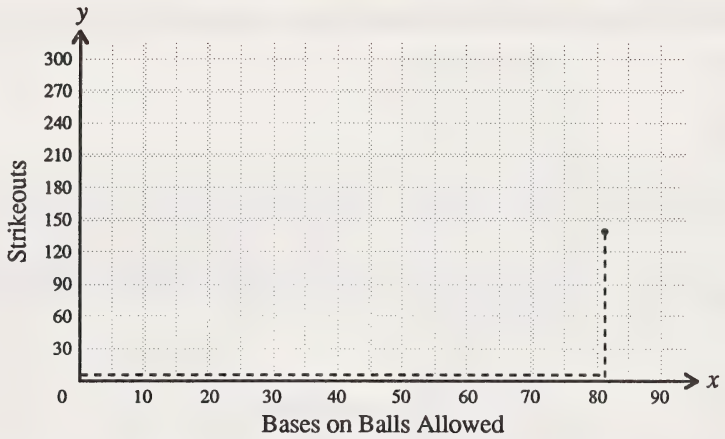
For this data, the number of bases on balls allowed will be placed along the horizontal axis and the number of strikeouts will be placed along the vertical axis. While examining the data, you will notice that the greatest number of bases on balls allowed by a pitcher was 89 and the greatest number of strikeouts was 298. You will have to make sure that both of these figures will fit onto the graph. Therefore, the horizontal axis will go to 90 and the vertical axis will go to 300. The graph used for the scatterplot is shown.



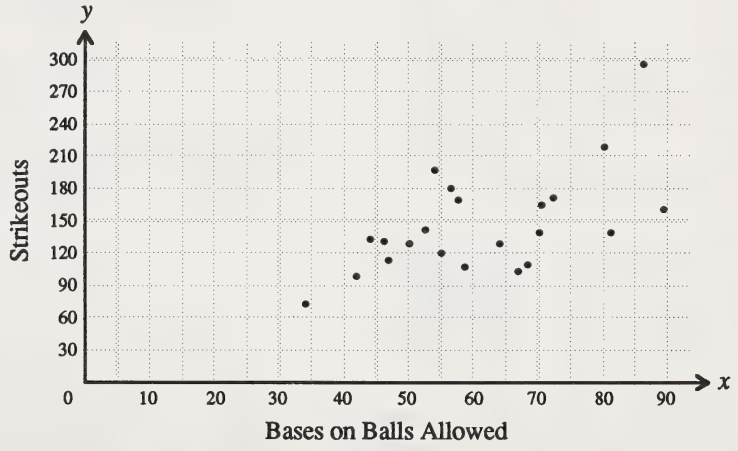
Step 2: Plot the data onto the scatterplot.

The data is plotted in exactly the same way you plot ordered pairs. The bases on balls allowed will be treated as the x-coordinate and the strikeouts will be treated as the y-coordinate.

To graph the point that represents Maddon, move from the origin 81 to the right and 140 up. This is the place where you will plot the first point. This movement is shown on the following graph.



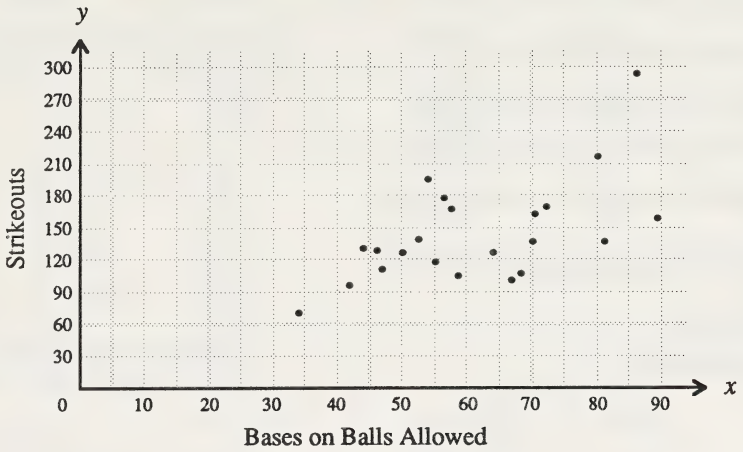
The rest of the points are plotted in the same manner. The scatterplot with all of the points plotted is shown.



Step 3: Name the scatterplot and label the axes.

Find a suitable name for the scatterplot.

Bases on Balls versus Strikeouts by Distance League Pitchers, 1989



The scatterplot is now complete.

While examining this graph, you will notice a trend. The more bases on balls allowed by a pitcher, the more likely it is that the pitcher is going to have more strikeouts. You should note that an increase in one of these two characteristics will not cause an increase in the other characteristic. An increase in both of these characteristics will depend on an increase in a third characteristic, namely, how many batters the pitcher has faced.

Since these two characteristics do not depend on each other, it did not matter on which axis each characteristic was placed. It would have been just as correct to place the bases on balls allowed along the vertical axis and the number of strikeouts along the horizontal axis.

In correlation problems it is not necessary to consider one variable to be dependent and the other to be independent.

However, there are times when the two characteristics must be placed along a certain axis.

Whenever one characteristic is dependent upon the other characteristic, the following rule should be observed.

The independent characteristic is always placed along the horizontal axis (or the x-axis), and the dependent characteristic is always placed along the vertical axis (or the y-axis).

The following example involves at a case where there are independent and dependent characteristics.

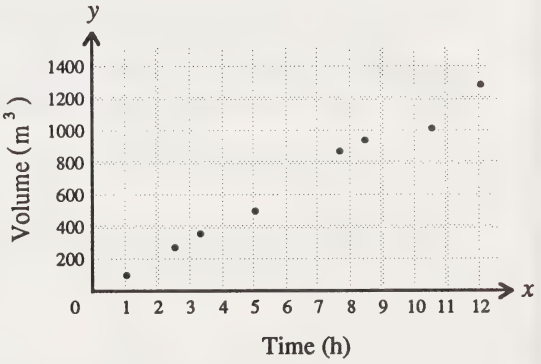
The following chart shows the amount of gas that has been transported through a pipeline after different lengths of time.

Time (h)	Volume (m^3)	Time (h)	Volume (m^3)
1	110	2.5	273
5	498	3.25	354
10.5	1056	8.4	915
12	1280	7.75	846

In this particular case the volume that has been transported through the pipeline is dependent upon the amount of time the gas has been flowing.

Therefore, time (independent characteristic) will be placed along the horizontal axis and volume (dependent characteristic) will be placed along the vertical axis.

Volume of Gas Transported Through Pipeline Omega
(January 1989)



Now you are ready to try some questions.

Do any three of the following questions.

1. For each of the following, identify whether the characteristics are independent, dependent, or have no relation.

   a. the mass and volume of water in a beaker

   b. the number of goals and assists scored by hockey players

   c. the distance travelled after a given amount of time

2. Make a scatterplot for the following information.

The Number of Wins and the Earned Run Average of Starting Pitchers in the Distance Baseball League, 1989

Pitcher	Wins	ERA
Maddon, Calgary	18	3.18
Prune, Edmonton	12	2.44
Conic, Lethbridge	20	2.22
Daze, Medicine Hat	13	3.67
Garney, Grande Prairie	12	3.69
Drysday, High Level	15	3.08
Shoulder, Lloydminster	16	3.26
Jakson, Drumheller	23	2.73
Reussel, Red Deer	19	3.12
Hornblower, Jasper	23	2.26
Scottie, Banff	14	2.92
Suttuer, Calgary	13	3.86
Martten, Edmonton	15	2.72
Coodie, Lethbridge	18	3.19
Terence, Medicine Hat	9	2.92
Carrie, Grande Prairie	10	4.29
Smilie, Stettler	13	3.25
Mussen, St. Paul	16	3.43
Browner, St. Albert	18	3.41
Drowns, Wetaskiwin	13	3.32
Larry, Banff	17	2.91
Knocks, Red Deer	14	3.14

3. Make a scatterplot for the following information.

The Mass and Fuel Consumption of Vehicles

Mass (kg)	Fuel Consumption (L/100 km)	Mass (kg)	Fuel Consumption (L/100 km)
800	5.0	1100	7.5
1300	7.8	1800	12.3
900	6.2	1400	10.4
1100	8.2	1700	12.0
1200	9.6	1000	8.5
600	3.9	1200	8.2
1500	9.8	900	5.6

4. Make a scatterplot for the following information.

Federal and Provincial Income Tax Payable for Northland Before Surtax, 1989

Federal	Provincial	Federal	Provincial
820	381.20	1060	493.60
1540	716.40	1780	828.80
2040	949.90	2280	1060.30
2560	1190.10	2620	1218.20
2740	1274.40	2860	1330.60
2980	1386.80	3020	1404.20
3160	1469.10	3340	1553.40
3660	1702.60	4240	1972.90

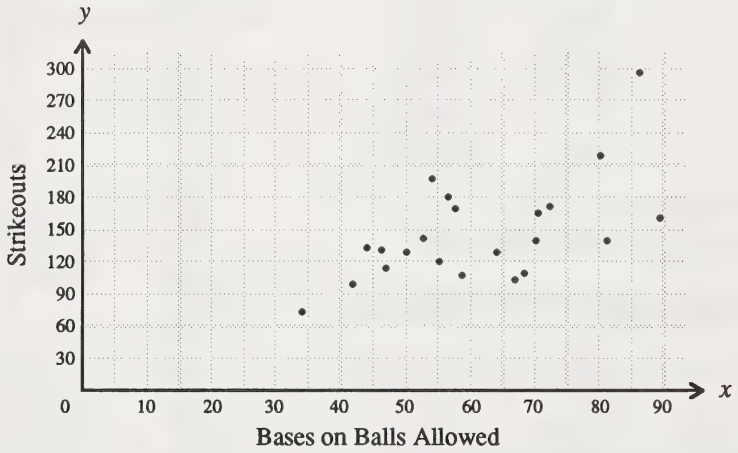
For solutions to **Activity 1**, turn to **Appendix A, Topic 1**.

Activity 2

Plot a line of best fit on a scatterplot using the median fit method.

Take another look at the pitcher scatterplot that was drawn in the first activity.

Bases on Balls versus Strikeouts by Distance League Pitchers, 1989



It is clear from the scatterplot that as the bases on balls allowed increase, so do the strikeouts. Unfortunately, the points on the scatterplot do not form a straight line.

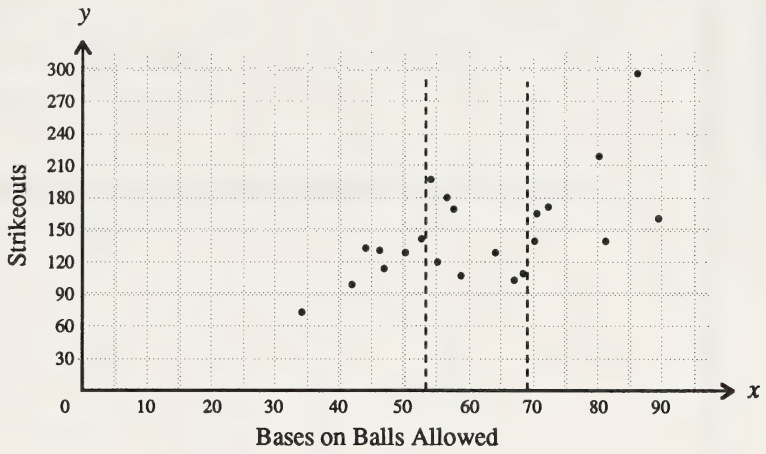
An approximation of the points can be formed to make a straight line. This approximation is referred to as **the line of best fit**.

The **median fit method** will be used to find the line of best fit.

Use this example to find the line of best fit using the median fit method.

Step 1: The points on the scatterplot are grouped into three vertical strips that have the same number of points. (If it is not possible for all three strips to have the same number of points, make sure the two outside strips have the same number of points.)

Bases on Balls versus Strikeouts by Distance League Pitchers, 1989



In this particular case there are twenty-two points. The points are divided into three strips so there are seven points in the first and last strips and eight points in the middle strip.

Step 2: The median point is found for each strip.

The first strip has the following points:

$(34, 65)$, $(42, 92)$, $(44, 131)$, $(46, 129)$, $(47, 118)$, $(50, 127)$, and $(53, 144)$

The median x-coordinate is as follows:

34, 42, 44, 46, 47, 50, 53

$\rightarrow 46$

The median y-coordinate is as follows:

65, 92, 118, 127, 129, 131, 144

$\rightarrow 127$

The median is the middle value from the data when the data is arranged in ascending order.

The second strip has the following points:

$(54, 190)$, $(55, 120)$, $(56, 180)$, $(57, 175)$, $(58, 112)$, $(64, 124)$, $(67, 103)$, and $(68, 116)$

The median x-coordinate is as follows:

54, 55, 56, 57, 58, 64, 67, 68

$\rightarrow \frac{57+58}{2} = 57.5$

The median of a group of data that has an even number of values is the average of the two middle values.

The median y-coordinate is as follows:

103, 112, 116, 120, 124, 175, 180, 190

$\rightarrow \frac{120+124}{2} = 122$

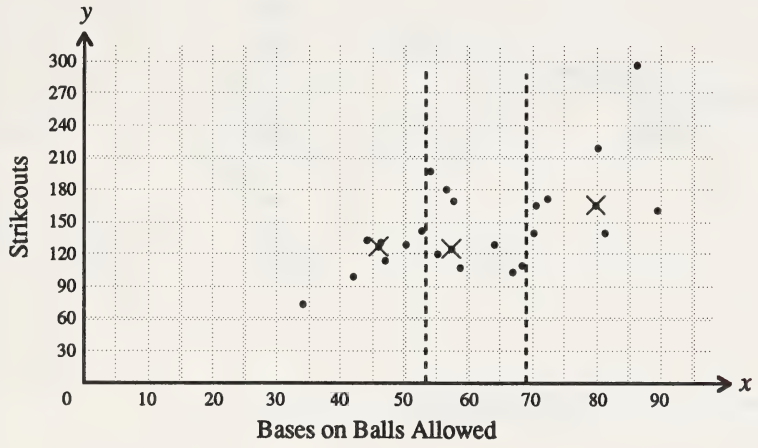
The third strip has the following points:
(70,144), (71,161), (73,178), (80,213), (81,140), (86,298), and (89,162)

The median x-coordinate is as follows:
70, 71, 73, 80, 81, 86, 89
→ 80

The median y-coordinate is as follows:
140, 144, 161, 162, 178, 213, 298
→ 162

Step 3: These points are then plotted onto the scatterplot.

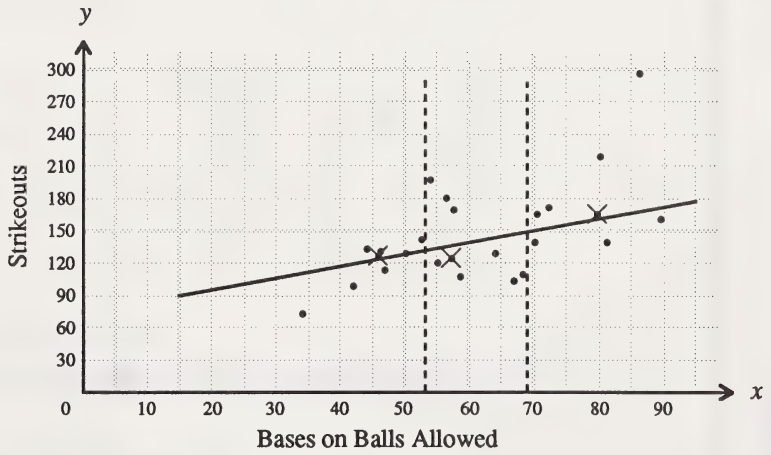
Bases on Balls versus Strikeouts by Distance League Pitchers, 1989



The median point can also be found by visual inspection. See the **Extra Help** section at the end of the topic.

Step 4: Place a ruler on the median points of the two outside strips. Slide the ruler, parallel to itself, one-third of the distance to the median point in the middle strip. Draw a straight line. This is the line of best fit.

Bases on Balls versus Strikeouts by Distance League Pitchers, 1989

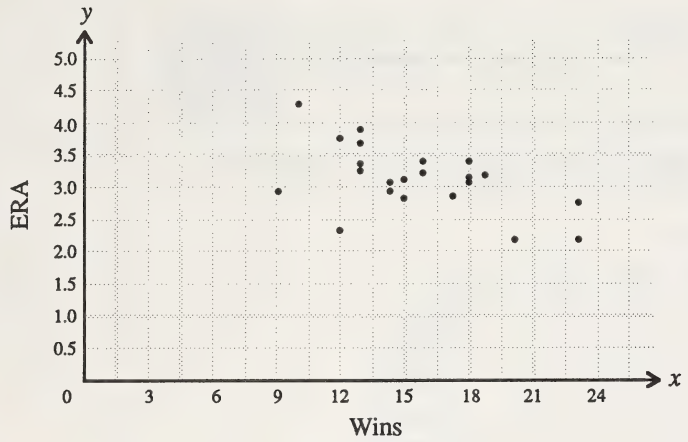


This line represents an approximation of the relation between the number of bases on balls allowed and the number of strikeouts by pitchers in the Distance Baseball League.

Do any two of the following questions.

1. Draw the line of best fit for the following scatterplot.

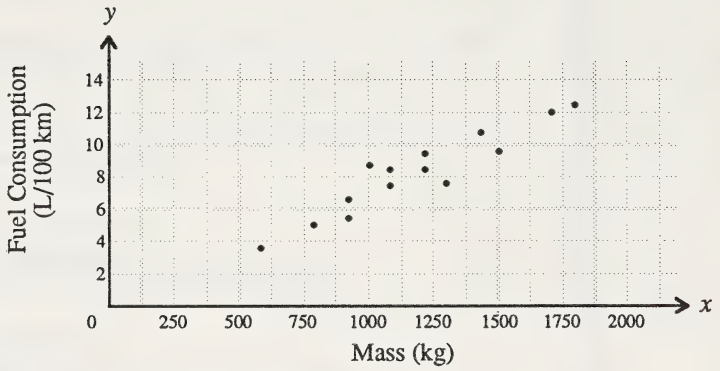
Wins versus ERA for Distance League Starting Pitchers, 1989
(ERA means earned run average.)



The coordinates of the points are given in Activity 1, question 2.

2. Draw the line of best fit for the following scatterplot.

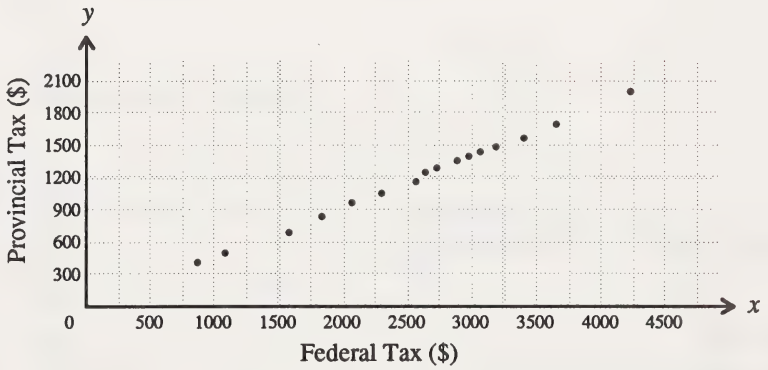
Mass versus Fuel Consumption of Vehicles



The coordinates of the points are given in Activity 1, question 3.

3. Draw the line of best fit for the following scatterplot.

Federal and Provincial Income Tax Payable in Northland
Before Surtax, 1989



The coordinates of the points are given in Activity 1, question 4.

For solutions to **Activity 2**, turn to **Appendix A, Topic 1**.

Activity 3

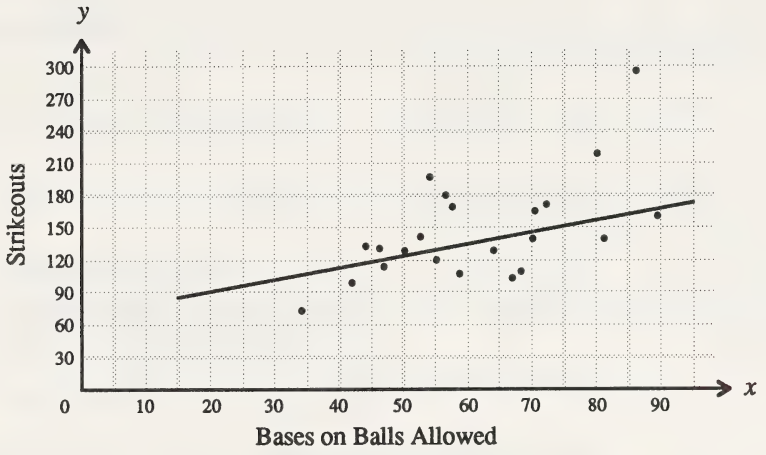
Use the equation of the line of best fit to generate new data for a population.

At this point you are able to take data, plot it on a scatterplot, and draw its line of best fit. Now you are ready to extend the width of the sample and find possible new values of data for the situation.

The key for extending the range of the data lies in the line of best fit. It is along this line that you will expect to find new pieces of data.

Take another look at the scatterplot that was made in the previous activity.

Bases on Balls versus Strikeouts by Distance League Pitchers, 1989

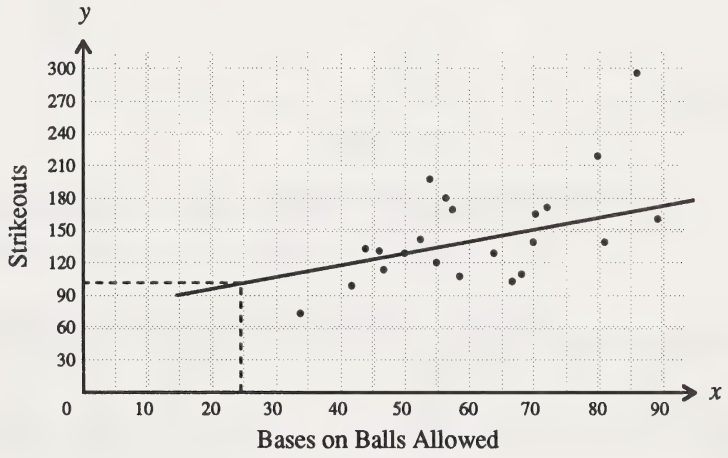


If a pitcher allowed twenty-five bases on balls, how many strikeouts would you expect the pitcher to have?

To find the solution to this question, you would go through the following steps:

1. Find 25 along the horizontal axis (or x-axis).
2. Move straight up to the line of best fit.
3. Move horizontally to the vertical axis.
4. Read the value from the vertical axis. This will be the number of strikeouts that you would expect the pitcher to obtain.

Bases on Balls versus Strikeouts by Distance League Pitchers, 1989



The preceding scatterplot shows that the pitcher would get approximately 100 strikeouts.

What happens when one of the values is not on the graph? Take a look at a particular case in which this happens.

How many bases on balls would you expect a pitcher to allow if that pitcher threw 240 strikeouts?

In this case you would have to start along the vertical or y-axis and move horizontally to the line of best fit. This is not possible with this scatterplot.

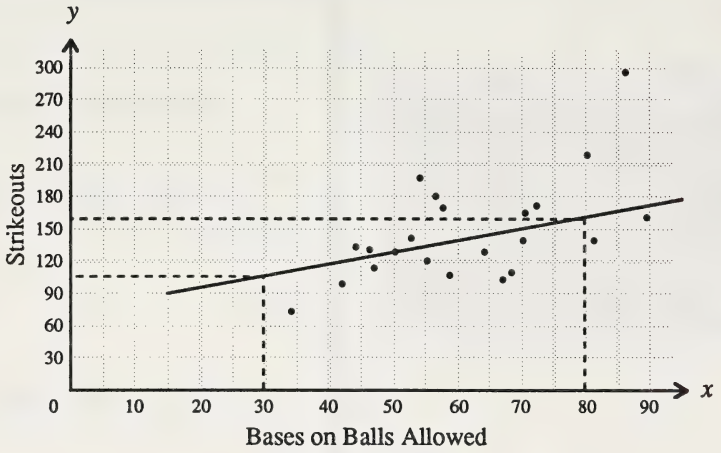
What you are looking for is a point along the line of best fit that has a y-coordinate of 240. This cannot be done visually.

If you find the equation of the line, you can solve for the missing coordinate algebraically.

Step 1: Find the equation of the line of best fit.

- First find the coordinates of any two points that are on the line.

Bases on Balls versus Strikeouts by Distance League Pitchers, 1989



The two points selected are $(30, 105)$, and $(80, 155)$.

- Next find the slope of the line.

$$m = \frac{y_2 - y_1}{x_2 - x_1}$$

$$= \frac{155 - 105}{80 - 30}$$

$$= \frac{50}{50} \text{ or } 1$$

The slope of the line is 1.

- Then solve for the y-intercept using the slope and y-intercept form of the equation.

$y = mx + b$

$y = 1x + b$ or

$y = x + b$

Substitute the point $(30, 105)$ into the equation and find the y-intercept.

$y = x + b$

$105 = 30 + b$

$b = 75$

Substituting the y-intercept into the equation will give you the equation of the line of best fit.

$y = x + b$

$y = x + 75$

The equation of the line of best fit is $y = x + 75$.

Step 2: Substitute the known coordinate into the equation of the line of best fit and solve for the missing coordinate.

$y = x + 75$

$240 = x + 75$

$x = 165$

This coordinate approximately represents the number of bases on balls allowed by a pitcher who throws 240 strikeouts.

This method will allow you to find new data (ordered pairs) anywhere along the line of best fit.

The point slope form can also be used to find the equation of the line.

$$y - y_1 = \frac{y_2 - y_1}{x_2 - x_1}(x - x_1)$$

$$y - 105 = \frac{155 - 105}{80 - 30}(x - 30)$$

$$y - 105 = \frac{50}{50}(x - 30)$$

$$y - 105 = x - 30$$

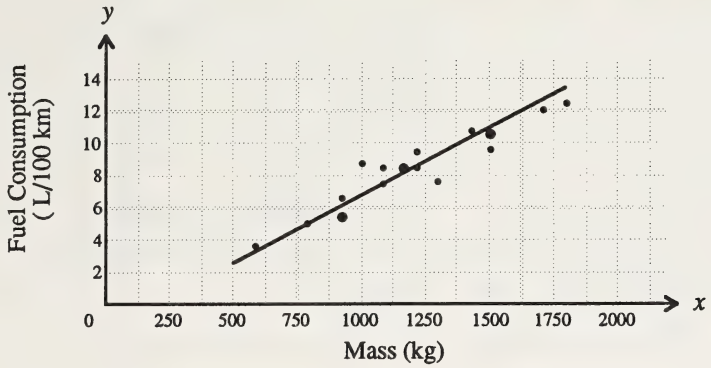
$$y = x + 75$$

This method is more accurate than estimating the missing coordinate from a scatterplot. Use this method whenever a more accurate answer is desired or when you are given two ordered pairs on the line of best fit.

Answer any two of the following questions.

1. Find the following missing values using the given scatterplot.

Mass versus Fuel Consumption of Vehicles



a. What is the fuel consumption for vehicles that have the following masses?

   i. 750 kg

   ii. 1275 kg

   iii. 1750 kg

   iv. 400 kg

   v. 4500 kg

b. What is the mass of each vehicle given the following amounts of fuel consumption?

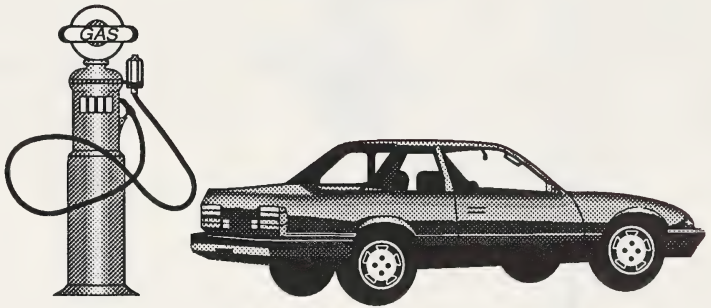
   i. 4.2 L/100 km

   ii. 8.8 L/100 km

   iii. 12.5 L/100 km

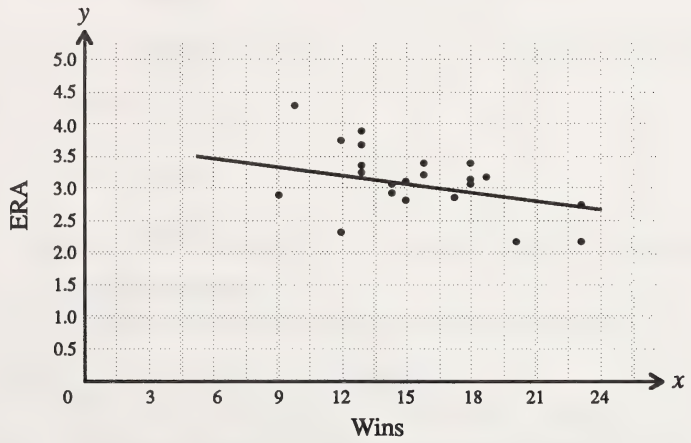
   iv. 1.2 L/100 km

   v. 17.9 L/100 km



2. Find the following missing values using the given scatterplot.

Wins versus ERA for Distance League Starting Pitchers, 1989
(ERA means earned run average.)



a. Find the number of wins a pitcher will have if the pitcher has an ERA of the following:

i. 3.40

ii. 3.20

iii. 3.00

iv. 3.60

v. 1.90

b. Find a pitcher's ERA if the pitcher has the following numbers of wins:

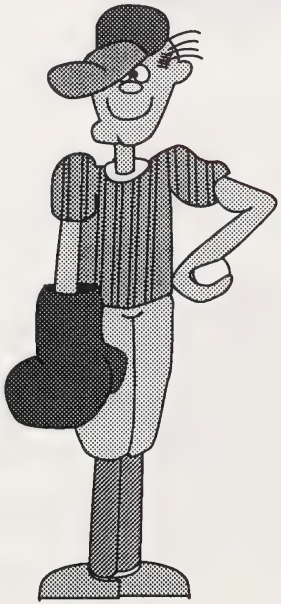
i. 6

ii. 18

iii. 22

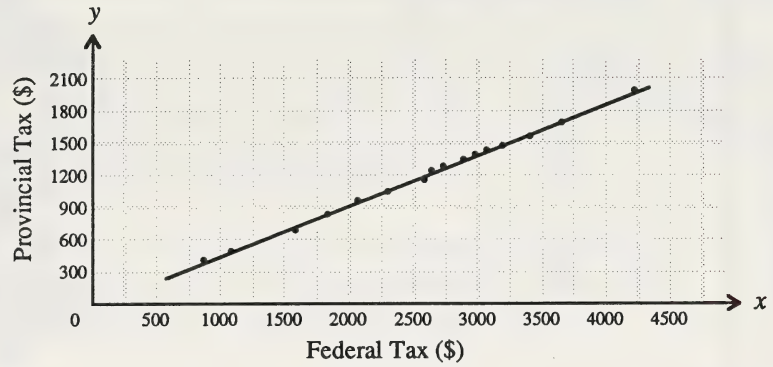
iv. 1

v. 30



3. Find the missing values using the following scatterplot.

Federal and Provincial Income Tax Payable in Northland
Before Surtax, 1989



a. What will be the federal tax when the provincial tax is the following?

i. $600

ii. $900

iii. $1800

iv. $2200

v. $3500

b. What will be the provincial tax when the federal tax is the following?

i. $1000

ii. $2500

iii. $3700

iv. $300

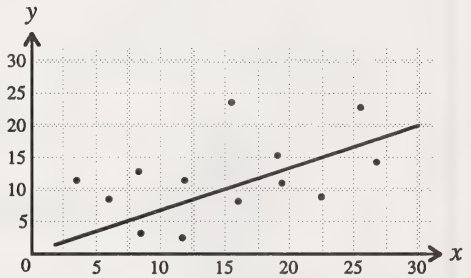
v. $5500

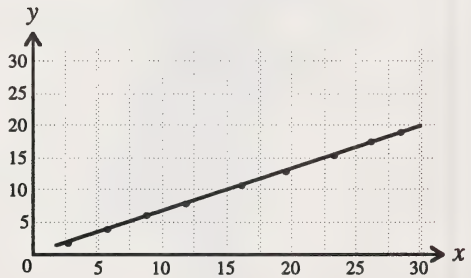
For solutions to **Activity 3**, turn to **Appendix A, Topic 1**.

Activity 4

Determine the strength and type of correlation between the variables of a bivariate distribution.

Examine the lines of best fit for the following two scatterplots.



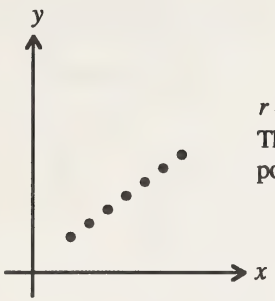


These two sets of data have the same line of best fit. However, the location of the data is very different in the two scatterplots. The data in the second scatterplot is much closer to the line of best fit.

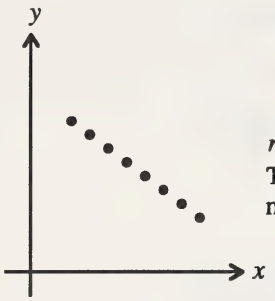
By examining the two scatterplots, you can also see that any data that is generated for the second scatterplot will be more reliable than the data generated for the first scatterplot. What you need to know is how much more reliable it will be.

What is needed is a way to tell how closely correlated the data is to the line of best fit. A number that lies between 1 and -1 is used to give a measure of how close the data is correlated to the line of best fit. This number is called the **correlation coefficient** (r). The correlation coefficient is not related to the numerical value of the slope of the line of best fit. You use the slope only to determine the sign of the correlation coefficient. The calculation of the correlation coefficient is beyond the scope of this course. Only the strength and type of correlation will be discussed here. (The calculation of the correlation coefficient is discussed in the **Extensions** section.)

If there are two variables x and y where all the ordered pairs (x, y) lie in a straight line and the slope of the line is positive, then the correlation coefficient(r) is 1. If the slope of the line is negative, then the correlation coefficient (r) is -1. The following diagrams show the two ideal situations.

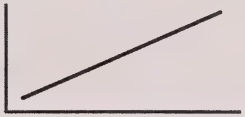


$r = +1$
This is called perfect positive correlation.

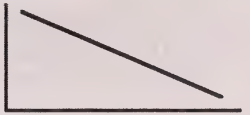


$r = -1$
This is called perfect negative correlation.

Perfect Positive Correlation

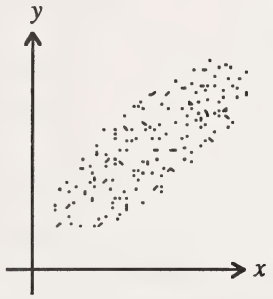


Perfect Negative Correlation

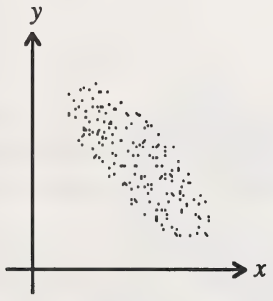


No Correlation

However, most of the time points do not lie on the same line. In such cases $|r|$ is between zero and 1. Here are some examples.



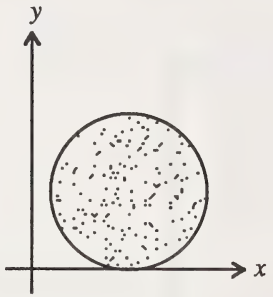
$r \doteq +0.7$



$r \doteq -\frac{1}{2}$

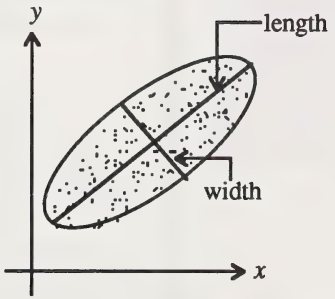
If the points are scattered farther away from the line of best fit, then the absolute value of r is smaller. No absolute value of r is greater than one.

If the points are scattered evenly all over the place or if they can be enclosed in a circle, then the correlation coefficient is zero.



$r = 0$

It is possible to estimate the correlation coefficient very roughly using a scatterplot. If you enclose the cloud of points in an ellipse, then the absolute value of the correlation coefficient is approximately equal to $\left(1 - \frac{\text{width}}{\text{length}}\right)$ and the slope of the line of best fit will determine the sign of the r.



For example, if the length of the ellispe is 5 and width is 2, then r is determined as follows:

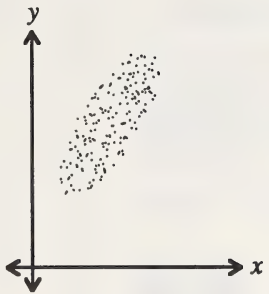
$$r = \left(1 - \frac{2}{5}\right)$$
$$= \frac{3}{5}$$
$$= 0.6$$

The sign r is positive because the slope of the line of best fit is positive. Again, this a very rough estimate. Students interested in learning how to calculate the correlation coefficient can study the **Extensions** section of this topic.

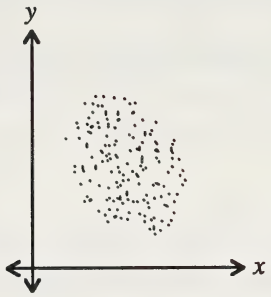
Now try the following questions.

Do questions 1 to 3.

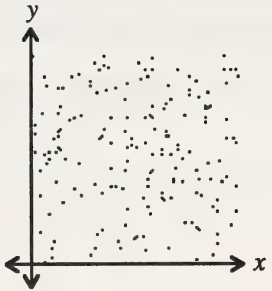
1. Estimate r.



2. Estimate r.



3. Estimate r.



For solutions to **Activtiy 4**, turn to **Appendix A, Topic 1**.

Activity 5

Collect, organize, and analyze sets of bivariate data.

Statistics can be viewed as a process that follows five basic steps.

- collection of data
- organization and presentation of data
- analysis of data
- drawing inferences
- evaluation for confidence

Look at how these steps are used in an actual situation.

Mr. Grant wanted to know how students' work habits influenced their grades.

Throughout the year, Mr. Grant gave the students two types of marks. The first mark showed the performance of the students on their work, assignments, projects, and tests. The second mark reflected the effort that the students put into their work. Mr. Grant thus completed the first of the five basic steps – collection of data.

Mr. Grant then organized the information into a chart. The chart had three columns: student identification, an average of the students' performance marks, and an average of the mark for student work habits. This information was then used to construct a scatterplot. In other words, Mr. Grant used the chart for organization and presentation of the data.

The line of best fit was found using the median fit method. Next, the equation of the line of best fit was found. This gave Mr. Grant a relationship between the performance and the work habits of the students. He could also use this equation to extrapolate new data. Thus, Mr. Grant's equation provided for an analysis of the data.

Mr. Grant then estimated the correlation coefficient. This gave Mr. Grant an indication of how well the line of best fit represented the data. Using this information, Mr. Grant was able to answer his original question. At this point, Mr. Grant would also analyze the situation and decide if one of the factors depends on the other factor and if there are any extraneous factors that influence both of these marks. At this point, therefore, Mr. Grant would be drawing inferences.

The evaluation for confidence will be discussed in the next topic.

You are now ready to do a project.

Perform the first four steps of statistics (collection of data, organization and presentation of data, analysis of data, and drawing inferences) on three of the following. Collect data from at least fifteen different people.

1. Compare the height of a person to the shoe size of that person.

2. Compare the height of a person to the waist measurement of that person.

3. Compare the height of a person to the finger stretch of that person. (Measure the finger stretch from finger tip to thumb tip across the back of the hand.)

4. Compare the height of a person to the hat size of that person. (For hat size, measure the circumference of the head.)

For a typical solution to the first project, turn to **Activity 5, Appendix A, Topic 1.**

If you require help, do the Extra Help section.

If you want more challenging explorations, do the Extensions section.

} You may decide to do both.

Extra Help

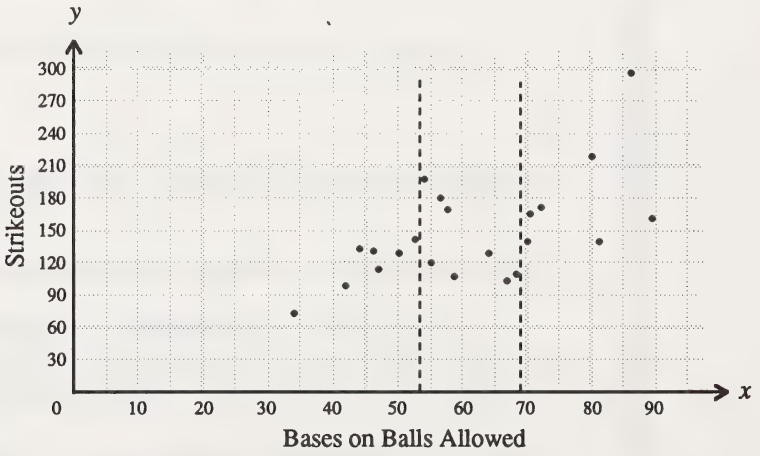
Finding the Line of Best Fit by the Inspection Method

Earlier in the topic, the median of each strip on the scatterplot was found by finding the median value of the x- and y-coordinates. This a long process. An alternate method is to find the median points by inspection.

The following are the steps that are used in finding the line of best fit using the inspection method.

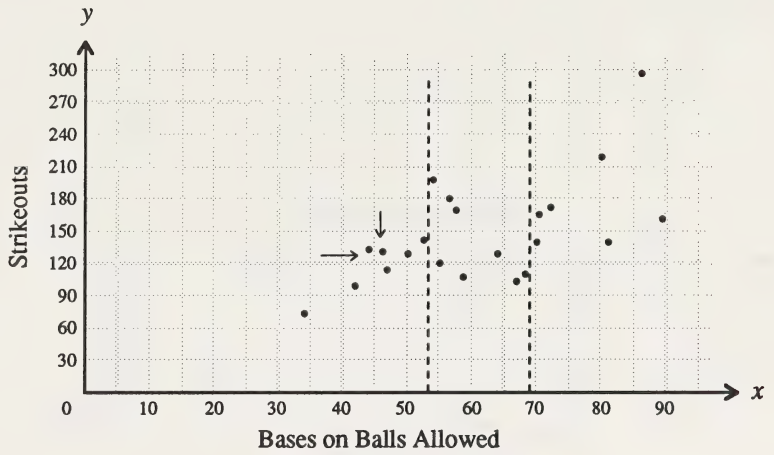
Step 1: Divide the scatterplot into three strips that have the same number of points in each strip. If it is not possible to put the same number of points in each strip, make sure that the first and last strips have the same number of points.

Bases on Balls versus Strikeouts by Distance League Pitchers, 1989



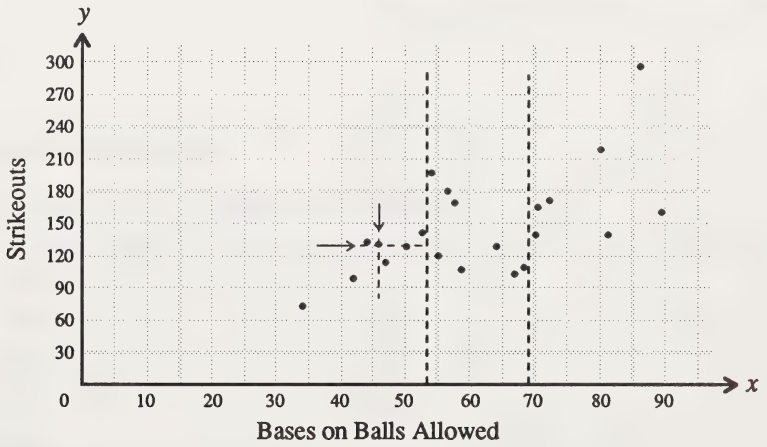
Step 2: Start with the first strip. If there are an odd number of points, select the middle point along the x-axis. If there are an even number of points, select a location halfway between the two middle points along the x-axis. Do the same for the y-axis.

Bases on Balls versus Strikeouts by Distance League Pitchers, 1989



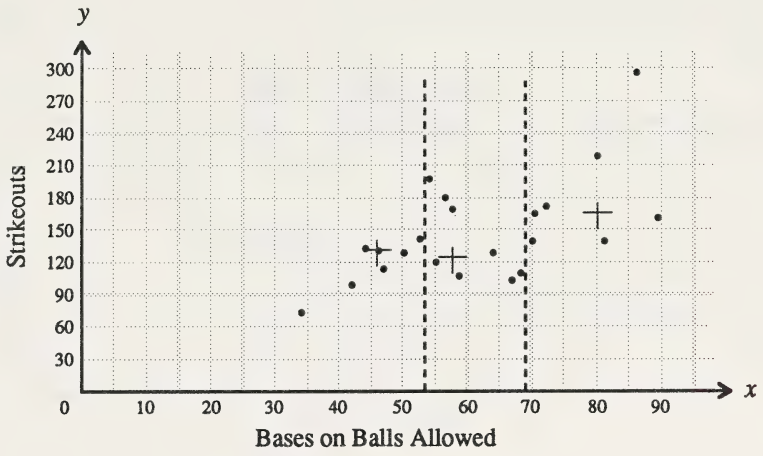
Step 3: Draw an imaginary line through the two arrows. The median point for the strip is located where the two lines intersect.

Bases on Balls versus Strikeouts by Distance League Pitchers, 1989



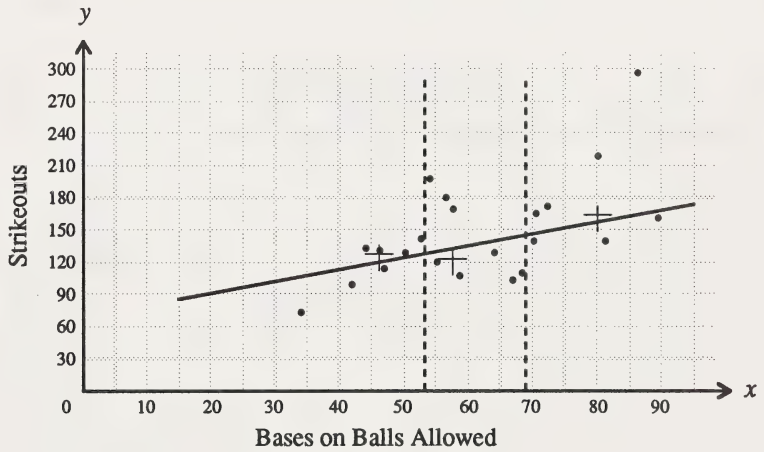
Step 4: Follow the same procedure as in Steps 2 and 3 and find the median points for the second and third strips.

Bases on Balls versus Strikeouts by Distance League Pitchers, 1989



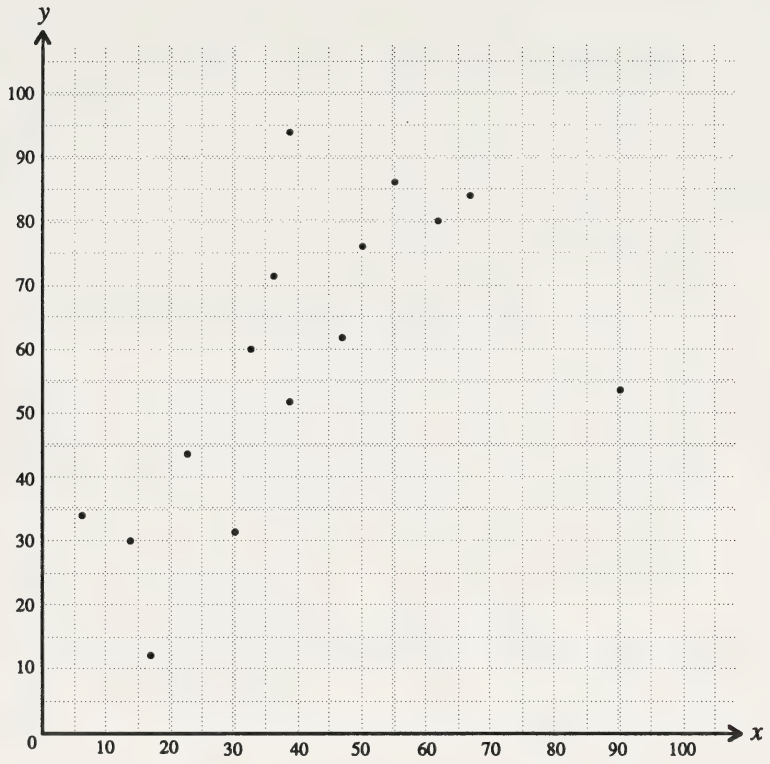
Step 5: The next step is to decide if the three median points lie on a straight line. You will notice if you use a ruler that these median points approximate a line. To determine the line, place your straightedge on the outside median points (median points in the first and last strips). Slide the straightedge a third of the distance to the middle median point (median point in the second strip) and draw the line. This will be the line of best fit.

Bases on Balls versus Strikeouts by Distance League Pitchers, 1989

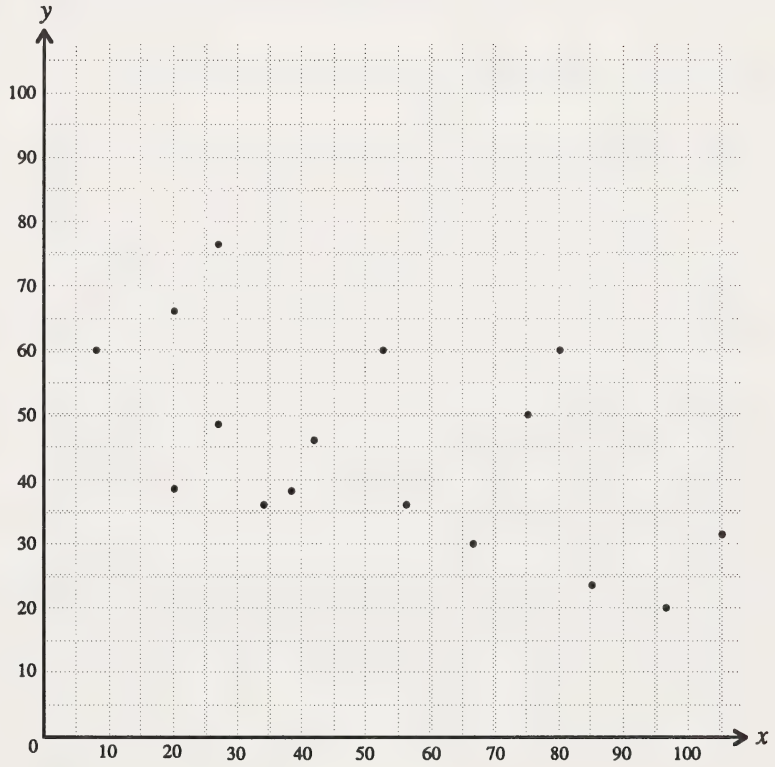


Try any two of the following questions.

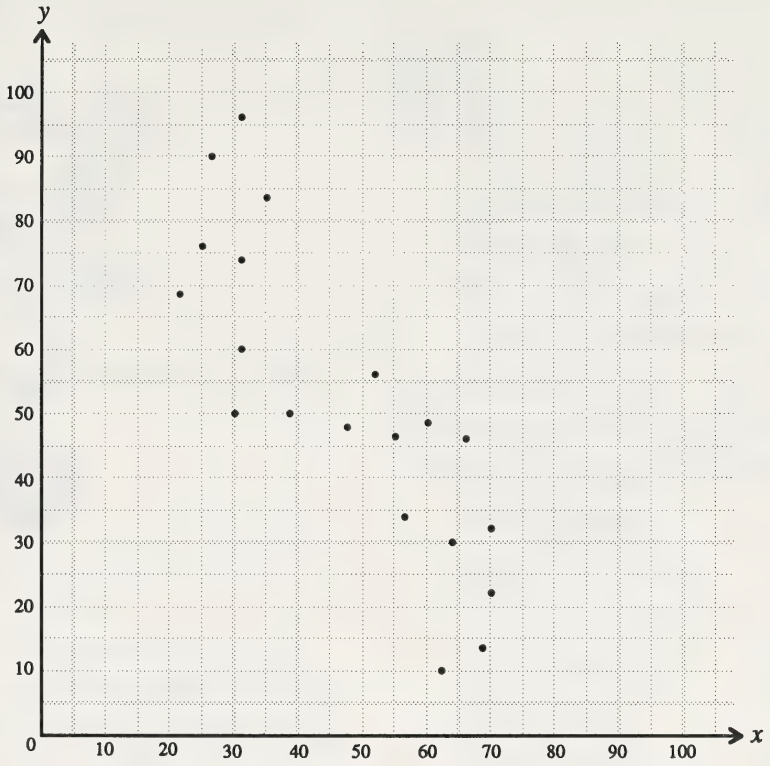
1. Use the inspection method to find the line of best fit for the following scatterplot.



2. Use the inspection method to find the line of best fit for the following scatterplot.



3. Use the inspection method to find the line of best fit for the following scatterplot.



For solutions to **Extra Help**, turn to **Appendix A, Topic 1.**

Extensions

If you have a graphics calculator it is very easy to calculate the correlation coefficient. The calculator used in this discussion is the Texas Instrument T1-81 which is available at any store which sells calculators. However, there are many different brand names. If you are using another graphics calculator, you will have to consult the manual yourself. Calculators are not designed the same way and it is impossible to show you how to use all of them to calculate the correlation coefficient.

Suppose that the bivariate data is given as follows:

Bases on Balls and Strikeouts by Pitchers in the Distance Baseball League, 1989

Pitcher	Bases on Balls	Strikeouts
Maddon, Calgary	81	140
Prune, Edmonton	44	131
Conic, Lethbridge	80	213
Daze, Medicine Hat	86	298
Garney, Grande Prairie	89	162
Drysday, High Level	50	127
Shoulder, Lloydminster	53	144
Jakson, Drumheller	71	161
Reussel, Red Deer	42	92
Hornblower, Jasper	73	178
Scottie, Banff	53	190
Suttuer, Calgary	70	140
Martten, Edmonton	55	120
Coodie, Lethbridge	57	175
Terence, Medicine Hat	34	65
Carrie, Grande Prairie	70	116
Smilie, Stettler	46	129
Mussen, St. Paul	58	112
Browner, St. Albert	64	124
Drowns, Wetaskiwin	47	118
Larry, Banff	56	180
Knocks, Red Deer	67	103

Step 1: Turn on the calculator and make sure that the calculator is reset to its factory settings. Clear everything in its memory. Adjust the display contrast to your liking.

Step 2: Press [2nd] [STAT].

Step 3: Move the cursor to DATA and press [1].

Step 4: Enter the number 81, then press [Enter].
The number 81 appears under x_1.

Step 5: Enter the number 140, then press [Enter].
The number 140 appears under y_1.

Step 6: Continue this procedure until you finish entering all values.

Step 7: Press [2nd] [STAT].

Step 8: Press [2].

Step 9: Press [Enter]. The screen displays a, b, and r. The correlation coefficient is r. In this case, $r = 0.603\,824\,064\,1$.

If you do not have a graphics calculator, the following formula can be used to calculate the correlation coefficient:

$$r_{xy} = \frac{\sum_{i=1}^{n}(x_i - \bar{x})(y_i - \bar{y})}{\sqrt{\sum_{i=1}^{n}(x_i - \bar{x})^2 \sum_{i=1}^{n}(y_i - \bar{y})^2}}$$

where x_i is each value

$\bar{x}$ is the mean

y_i is each value

$\bar{y}$ is the mean

This procedure is a long one. You may want to try the same question using this formula. Your answer should be very close to 0.603 824 064 1, but not exactly the same.

Now do one of the following questions using your graphics calculator.

1. Find the correlation coefficient for the following bivariate data.

The Number of Wins and Earned Run Average of Starting Pitchers in the Distance Baseball League, 1989

Pitcher	Wins	ERA
Maddon, Calgary	18	3.18
Prune, Edmonton	12	2.44
Conic, Lethbridge	20	2.22
Daze, Medicine Hat	13	3.67
Garney, Grande Prairie	12	3.69
Drysday, High Level	15	3.08
Shoulder, Lloydminster	16	3.26
Jakson, Drumheller	23	2.73
Reussel, Red Deer	19	3.12
Hornblower, Jasper	23	2.26
Scottie, Banff	14	2.92
Suttuer, Calgary	13	3.86
Martten, Edmonton	15	2.72
Coodie, Lethbridge	18	3.19
Terence, Medicine Hat	9	2.92
Carrie, Grande Prairie	10	4.29
Smilie, Stettler	13	3.25
Mussen, St. Paul	16	3.43
Browner, St. Albert	18	3.41
Drowns, Wetaskiwin	13	3.32
Larry, Banff	17	2.91

2. Find the correlation coefficient for the following bivariate data.

The Mass and Fuel Consumption of Vehicles

Mass (kg)	Fuel Consumption (litres/100 km)	Mass (kg)	Fuel Consumption (litres/100 km)
800	5.0	1100	7.5
1300	7.8	1800	12.3
900	6.2	1400	10.4
1100	8.2	1700	12.0
1200	9.6	1000	8.5
600	3.9	1200	8.2
1500	9.8	900	5.6

3. Find the correlation coefficient for the following bivariate data.

Federal and Provincial Income Tax Payable for Canuckland, Before Surtax 1989

Federal	Provincial	Federal	Provincial
820	381.20	1060	493.60
1540	716.40	1780	828.80
2040	949.90	2280	1060.30
2560	1190.10	2620	1218.20
2740	1274.40	2860	1330.60
2980	1386.80	3020	1404.20
3160	1469.10	3340	1553.40
3660	1702.60	4240	1972.90

For solutions to **Extensions**, turn to **Appendix A, Topic 1**.

Topic 2 Confidence

Introduction

With the use of statistics you are now able to draw more reliable conclusions and make more accurate inferences about different situations.

This is still not enough. You need to know how accurate your conclusions and inferences are before the event happens.

The accuracy of your conclusions and inferences is referred to as the confidence of your conclusions and inferences and this is what you will be doing in this topic.

What Lies Ahead

Throughout this topic you will learn to

1. design and administer a yes/no simple survey and collect and organize the results of the survey
2. draw box and whisker plots of the results of multiple samples
3. use a 90% box and whisker plot chart to find the confidence interval for a survey result
4. draw statistical conclusions and make inferences to populations, and explain the confidence with which such conclusions and inferences are made based on the results of yes/no surveys
5. assess the strengths, weaknesses, and biases of given samples

Now that you know what to expect, turn the page to begin your study of confidence.

Exploring Topic 2

Activity 1

Design and administer a yes/no simple survey and collect and organize the results of the survey.

The Canadian government through Statistics Canada conducts two different types of census. The decennial census (conducted once every ten years) was first conducted in 1851 and is responsible by law to provide population counts of electoral districts to enable any redistributions of the seats in the House of Commons.

The quinquennial census (conducted once every five years) was first conducted in 1956 and is used to keep the nation's statistical information abreast of the demographic and socioeconomic developments that affect decision making in both the public and private sectors.

These two censuses are considered necessary to the well-being of the citizens of the nation.

Through the census, the decision makers of the nation can effectively plan to meet the needs of the citizens by the development and implementation of policies and programs.

Both of these censuses were most recently conducted in 1991.

The undertaking of these censuses is truly mammoth. Every household in the nation will receive census forms to be filled out and returned to Statistics Canada. Once the forms are returned the long task of compiling the information begins.

The information that has been collected by the censuses and other surveys conducted by Statistics Canada can be found in the Canada Year Book which has been published annually since 1959.

Often information is needed more quickly than it can be delivered by a census. At these times, a **survey** is taken. An example of a survey is the statistics compiled to determine the unemployment rate each month.

Gathering information has become so important to society that you will find other groups such as television and radio stations and newspapers that conduct surveys and opinion polls for their listeners and readers.

Since the information gathered by these censuses is so important to proper decision making in Canada, every household is required by law to complete and return the forms.

There are firms such as Gallup, Angus Reid, and Environics which make profits from conducting surveys for other groups.

This topic is not going to deal with the statistics that arise from the census, but instead the statistics that permit these firms to make statements about an entire group of people, or **population**, after they have only questioned a small group of these people, or a **sample** of the population.

Surveys or polls usually ask more than one question, and each question has the possibility of more than two responses. To help with your understanding of this material, only single survey questions that have yes or no responses will be used. The following is one such question.

> Do you approve of the way Brian Mulroney is handling the Canadian government?

There are only two ways to answer this question: yes or no.

A third response of undecided could also be used here. This response will be avoided to simplify the calculations.

This question was asked by an imaginary pollster, Runners. Look at the article that gave the results of this poll.

> A survey in March of 1989 conducted by Runners asked 1257 adult Canadians the following question.
>
> "Do you approve of the way Brian Mulroney is handling the Canadian government?"
>
> Forty-three percent said that they approved. For results based on samples of this size, we are 95% confident that the error attributable to sampling and other random effects could be three percentage points above or below the given results.
>
> The reader should also remember that the wording of the question and practical difficulties encountered in conducting the survey can introduce errors or bias into the findings of the survey.

What information is given in this article?

- The survey was conducted in March of 1989.
- The survey included 1257 adults.
- Forty-three percent of the sample said yes to the question.
- If the entire population was asked, 40% to 46% of the people would have responded yes 95% of the time.
- There may have been unknown circumstances that biased the findings.

This is a large survey and the method used is different from what you will learn in this topic.

It is fairly simple to see where the first three conclusions came from, but how did the survey firm arrive at the 95% confidence and the three percentage points above and below 43%?

Runners feels confident, based on its calculations, that if the entire population was surveyed, between 40% and 46% of the people would say yes to the question (3% above and below 43%). This statement is called a **statistical inference**, a prediction based on the gathered and organized data.

The 40% to 46% range (which is found by adding three to 43% and subtracting three from 43%) is called the **confidence interval**. The **confidence interval** is the range in which most of the samples would fall. (In this case, 95 out of every 100 samples would fall within the range.)

Now consider a smaller survey and assume the **population percentage** is 43%. The **population percentage** will be the percentage of the survey sample that responded yes.

Suppose you asked thirty people the survey question. How many would say yes to the question?

To perform this survey, you will have to make sure that the people are selected randomly. This is quite a task. You will have to make sure that people from all parts of the country are just as likely to be selected. You will have to make sure that all adult age groups are just as likely to be selected. You will have to make sure that men and women are just as likely to be selected. Also, other criteria will have to be considered such as religion, political affiliation, and socioeconomic status.

Instead of going to all of the trouble of trying to meet these criteria, a simulation will be used to generate the responses.

There are a number of ways that you can simulate this experiment. You could use a random number table, a spinner, the random number generator of a microcomputer, or you could draw marbles from a vat.

The 3% is called a sampling error. This sampling error can be calculated by using the formula $2 \bullet \sqrt{\frac{P(1-P)}{n}}$, but you will not use this method at this time.

See the **Extensions** section at the end of this topic to see how to use a microcomputer to simulate these selections.

For this simulation, marbles will be drawn from a vat. The marbles will carry one of two labels, either yes or no.

To meet the conditions of the example, 43% of the marbles in the vat must have the label yes. Also, it will be necessary to make sure that the marbles are properly mixed and that they are the same size. Otherwise this will not be a random selection.

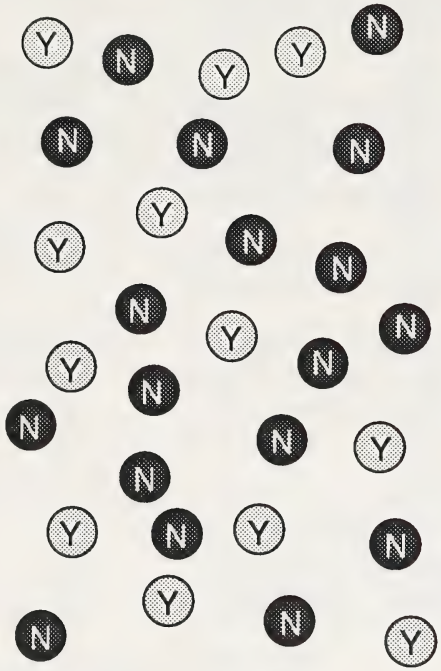
The population percentage can only be found by counting the number of yeses in the vat. If there are more than eight million marbles in the vat, it would be unreasonable to count all of the marbles. However, if a sample of the marbles is taken from the vat, the population percentage can be estimated by the **sample proportion**.

The sample proportion can be found by dividing the number of yeses in the sample by the sample size.

Thirty marbles are removed from the vat. Of the thirty marbles, twelve are labeled with the word yes. Therefore, the sample proportion is $\frac{12}{30}$ or 0.4.

This sample proportion represents 40%. If you returned the marbles to the vat, mixed them again, and then selected another thirty marbles, you would get another sample proportion.

Select thirty marbles from the vat forty times. Be sure that you replace the marbles and mix them in the vat after the results of each draw of thirty marbles are recorded.



The **population percentage** will be the percentage of the objects in the vat that are **yeses** (labeled with the word yes).

The following is an example of these results.

Experiment	Responses	Number of Yeses
1	NYYNNNNNYNNYNNNYYNYYNNNNYNYNNN	10
2	YNYYNYNNYYNYYYYNNNNNYYYYYNNYYY	18
3	NNYNYYYNYYNYNYYNYNNYNYYNYNYNN	16
4	NNYNNNYYYNYNYNNNNYNYYYYNNYNNNN	12
5	NYYNYYNNNNNYNNYYYYYYNNNNNNYNYY	14
6	NNNYNNNNNYYNNYYYNNNNNYYYNNNNNY	10
7	YYNYNNNNYNYNYYNYNNYNYNYYNNNNNY	13
8	YNNNYYNYYNNYYYYNNNYYYYNNYNNNN	14
9	NNNNNNYNYYYYYNNNYNNNYYYYNNYNNY	13
10	NNYNYNYNYNNYYYYYNNNNNYYYNNYNNN	13
11	YNNYYYYYYNNYNYYNNNYNNNNYNNYNYN	14
12	YNNYNNNNNNYNNNNYNNYNNYYYNNNYNN	9
13	YNNYNNNNYNYYYNNYNNYYYNYYNYYNYN	15
14	YNYYNYNYYYNNYNYNNYNNYYNYYNNNNY	15
15	YNNYNNYYYNYNYNYNNYNNNNNNNYYNNY	12
16	YNNNNYYNNNNYNNNNYYYNNNNNNYNYNN	9
17	NNYNYNNNNNNNNNYNNYNNYYYYYYNYNN	11
18	YNNNNYYNNYNNYNNYYNYNNYNNYYNN	12
19	NYYNYYNYNNNYNYNNYNNNNYNYNNNNNY	11
20	YYYYYYNYNNNYNNNYYYNNNYYNYYNNNY	16
21	NNYNNYNNNYYNYYYNNYYNYYNNYYNYNY	15
22	NYYNNNYNYYYNYYNNYNYYYNNNYYNNN	15
23	YYNNNYYYNYYYYYNYYNNYYNNNNNYYYY	18
24	YNNNNYNNNNYNYYYNYYNNNNYYYNYNN	13
25	NNYYNYNNYYYYYNNNYYNNNNYNNNYYNN	13
26	NYNYNNNYYNNNYYNNYYNNYYYNYNNNNY	13
27	YYYNYNNNNNNYYNYYNYNNYYYNNYYNNN	14
28	NNNNYNYNNYNYNNNYNNYNYYNNYNNYYN	11
29	NNYNYYNYNYNNNYNNNYYYYYNYYYYYYN	18
30	YYNNNNNNNYNNNNNYYYNYNNNYNNNNNN	8
31	NNNNYYNYYYNNNNNYYNNNNYYNYYNNYN	12
32	YNNYNNNNYYYNYYYNNNYNYYYNYNNNYN	14
33	YYYYYYNYNNNYNYYYYYNYNNNNNYYNYY	18
34	NNNNYNYNNNNYYNNYNYNNNNNNNNYYYN	9
35	NNYNNNNYYYYNYNYNYNNYYNNNYNNYYY	14
36	NYNNNNNYNNYYNYYNYNNNNYNYYNNYYN	12
37	YNYYNNNYYNYNNYYNYYYYYYNNYYNNYY	18
38	YNNNYYNYNNNYNYYYNYYNNYYNNYYYYY	17
39	YNNYYNNYNNNYYNNNNYNYYYYNYNYNYY	15
40	NNYNNYNYNNNYYNYNNYNYNNNNYNYYNYY	13

Once the data is collected, it should be sorted into a **sampling distribution.** The following sampling distribution has been collected from the table of forty groups of thirty marbles selected from the vat.

Number of Yeses	Sample Proportion	Tally	Frequency	Proportion of All Trials
0	0.00		0	0.000
1	0.03		0	0.000
2	0.07		0	0.000
3	0.10		0	0.000
4	0.13		0	0.000
5	0.17		0	0.000
6	0.20		0	0.000
7	0.23		0	0.000
8	0.27	I	1	0.025
9	0.30	III	3	0.075
10	0.33	II	2	0.050
11	0.37	III	3	0.075
12	0.40	IIIII	5	0.125
13	0.43	IIIII II	7	0.175
14	0.47	IIIII I	6	0.150
15	0.50	IIIII	5	0.125
16	0.53	II	2	0.050
17	0.57	I	1	0.025
18	0.60	IIIII	5	0.125
19	0.63		0	0.000
20	0.67		0	0.000
21	0.70		0	0.000
22	0.73		0	0.000
23	0.77		0	0.000
24	0.80		0	0.000
25	0.83		0	0.000
26	0.87		0	0.000
27	0.90		0	0.000
28	0.93		0	0.000
29	0.97		0	0.000
30	1.00		0	0.000
Total			40	1.000

Do you remember how the sampling distribution is constructed? If not, answer the following questions about the example results and the sampling distribution.

How many times were there ten yeses in the example results? What is the frequency for ten yeses from the sampling distribution? Both answers are 2.

How many times were there fifteen yeses in the example results? What is the frequency for fifteen yeses from the sampling distribution? Both answers are 5.

The frequency states the number of experiments that contain the specified number of yeses. For example, only one experiment contained seventeen yeses. Check the example results and sampling distribution to make sure that this is correct.

What is the most number of yeses that can be collected in each experiment? Since there are thirty responses for each experiment, the most number of yeses is 30.

What is 25 divided by 30 rounded to two decimal places? What is the sample proportion for twenty-five yeses from the sampling distribution? Both of these answers are 0.83.

What is 18 divided by 30? What is the sample proportion for eighteen yeses from the sampling distribution? Both of these answers are 0.60.

The sample proportion is the number of yeses divided by the total number of responses (or the highest possible number of yeses).

There was a total of forty experiments conducted for this example. The proportion of all trials is found by dividing the frequency by the total number of experiments, 40.

Examine this sampling distribution carefully. Does this distribution look familiar? The distribution resembles the frequency distribution table. This table keeps track of the number of samples that contained a specified number of yeses.

This sampling distribution is an approximation through simulation. The sampling distribution is the number of yeses in a sample of size 30.

This sampling distribution shows the amount of variability that exists between different random samples of a population. From this example, you would expect the number of yeses to be between 8 and 18 or 27% and 60%.

This confidence interval is much larger than the one given in the large survey of 1257 adults. Why?

This is not the exact sampling distribution for the large survey. The sample should have a size of 1257. If you selected samples as large as those given in the example (1257), you are more likely to get the same confidence interval.

In general, to construct a sampling distribution, both the population percentage and the sample size must be known. With the sampling distribution, you can use the number of yeses in the sample or its equivalent form, the sample proportion, to determine the probability of getting a certain number of yeses or a certain sample proportion.

Do any three of the following questions.

1. Write survey questions to find the following information. The questions should only ask for yes and no responses.

   a. Do people prefer to watch hockey or baseball?

   b. Do people prefer large or small breakfasts?

   c. Do people prefer small or large cars?

   d. Do people prefer to vacation in Europe or in the sun on a Pacific island?

2. Write a series of yes/no survey questions that will find the following information.

   a. Which of the following fruits do people prefer to eat: bananas, oranges, or apples?

   b. For which of the following politicians would people vote to become prime minister: Brian Mulroney, Audrey McLaughlin, or Jean Chrétien?

3. Toss eight coins one hundred times. Record the results in a sampling distribution. Use your results to answer the following questions.

   a. What is the most likely number of heads each time you toss the eight coins?

   b. Estimate the probability of getting 2, 3, 7, and 8 heads? (Give four answers.)

   c. Estimate the probability of getting 2, 3, 7, or 8 heads? (Give one answer.)

4. Explain how you would set up a simulation for the following situation:

   Out of a sample of twenty people, fourteen people prefer to vacation in the Caribbean.

For solutions to **Activity 1**, turn to **Appendix A, Topic 2**.

Activity 2

Draw box and whisker plots of the results of multiple samples.

In the last activity you constructed a sampling distribution. In this activity you are going to take that sampling distribution and use the information to construct a 90% box and whisker plot.

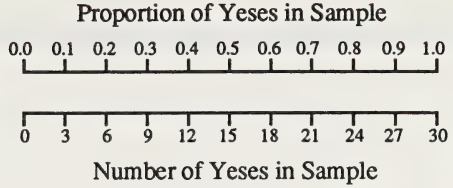
The sampling distribution that you constructed in Activity 1 came from the following survey question:

> Do you approve of the way Brian Mulroney is handling the Canadian government?

This same sampling distribution is shown.

The following steps are used to construct a box and whisker plot for the sampling distribution.

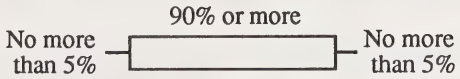
Step 1: Construct two axes. The first one is for the proportion of yeses in the sample, and the second one is for the number of yeses in the sample.



Number of Yeses	Sample Proportion	Tally	Frequency	Proportion of All Trials
0	0.00		0	0.000
1	0.03		0	0.000
2	0.07		0	0.000
3	0.10		0	0.000
4	0.13		0	0.000
5	0.17		0	0.000
6	0.20		0	0.000
7	0.23		0	0.000
8	0.27	I	1	0.025
9	0.30	III	3	0.075
10	0.33	II	2	0.050
11	0.37	III	3	0.075
12	0.40	IIIII	5	0.125
13	0.43	IIIII II	7	0.175
14	0.47	IIIII I	6	0.150
15	0.50	IIIII	5	0.125
16	0.53	II	2	0.050
17	0.57	I	1	0.025
18	0.60	IIIII	5	0.125
19	0.63		0	0.000
20	0.67		0	0.000
21	0.70		0	0.000
22	0.73		0	0.000
23	0.77		0	0.000
24	0.80		0	0.000
25	0.83		0	0.000
26	0.87		0	0.000
27	0.90		0	0.000
28	0.93		0	0.000
29	0.97		0	0.000
30	1.00		0	0.000
Total			40	1.000

Most surveys use a 95% box and whisker plot. The 90% box and whisker plot is done here since it is easier to calculate.

Step 2: Find the number of values that will be in the box and the whiskers. Since you are making a 90% box and whisker plot, 90% of the values must be in the box and 5% of the values will be in each whisker. The box can have more than 90% of the values, but any one whisker will have at most 5% of the values.



In this particular case there are forty trials. There will be
90% of $40 = 0.9 \times 40 = 36$
values in the box. Also, there will be
5% of $40 = 0.05 \times 40 = 2$
values in each of the whiskers.

Step 3: Find the number of yeses for these locations. The completed diagram for this box and whisker plot is in Step 5, so refer to it as you read these instructions for Step 3.

The location of the left endpoint for the left whisker is 8 or 0.27. See the diagram in Step 5.

The location of the point where the left whisker joins the box is 9 or 0.3. The left end of the box starts at 9 or 0.3 since the frequency for nine yeses is 3 and the left whisker can have no more than two values. The left whisker already has 1 for the frequency which represents the eight yeses.

The location of the point where the right whisker joins the right end of the box is 18 or 0.6. The reason is that the frequency is 5 for the eighteen yeses and these five cannot be separated. Thus, the box contains thirty-nine values. This means that there is no right whisker on the diagram. The location of the right endpoint for the right whisker can be interpreted as being at 18 or 0.6.

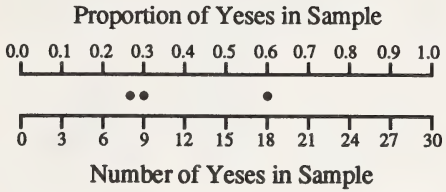
As you can see from these results, the endpoint for the right whisker and the end of the right side of the box will be the same point.

For a 95% box and whisker plot, you put 95% of the values in the box and 2.5% of the values into each whisker.

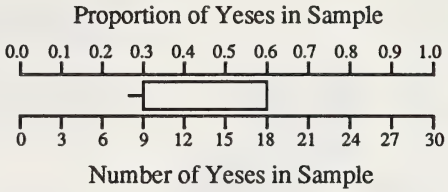
Because of these ties in the sample proportions, you cannot get exactly thirty-six values in the box. This box actually contains thirty-nine values, counting the edges of the box. When a tie occurs, such as here, you will always construct the plot so that no more than 5% of the values are in either whisker. The box can have more than 90% of the values, but any one whisker will have at most 5% of the values.

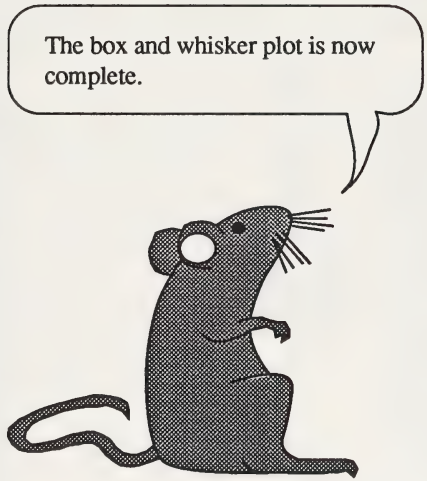
The case you are doing now is a good example of this. There will be only one point outside of the box on the left side, which is 2.5% of the values. Also, there will be no values outside the box on the right side, or 0% of the values. This means that there is 97.5% of the values in the box.

Step 4: Find the ends of the box and the whiskers on the axes. Mark these locations with dots.



Step 5: Using the dots as your guides, draw in the box and the whiskers.





If you put the left edge of the box at 10 instead of 9, you would have four values in the whisker. This is unacceptable since you are allowed only two values in the left whisker.

If you would like more work on creating 90% box and whisker plots, go to the **Extra Help** section at the end of the topic.

Do either one of the following questions.

1. Make a 90% box and whisker plot for the following sampling distribution.

Number of Yeses	Sample Proportion	Tally	Frequency	Proportion of All Trials
0	0.00		0	0.000
1	0.05		0	0.000
2	0.10		0	0.000
3	0.15		0	0.000
4	0.20		0	0.000
5	0.25		0	0.000
6	0.30		0	0.000
7	0.35		0	0.000
8	0.40		0	0.000
9	0.45		1	0.017
10	0.50		5	0.083
11	0.55		11	0.183
12	0.60		13	0.217
13	0.65		12	0.200
14	0.70		10	0.167
15	0.75		6	0.100
16	0.80		2	0.033
17	0.85		0	0.000
18	0.90		0	0.000
19	0.95		0	0.000
20	1.00		0	0.000
Total			60	1.000

2. Make a 90% box and whisker plot for the following sampling distribution.

Number of Yeses	Sample Proportion	Tally	Frequency	Proportion of All Trials
0	0.00		0	0.00
1	0.05		0	0.00
2	0.10		0	0.00
3	0.15		0	0.00
4	0.20		0	0.00
5	0.25		0	0.00
6	0.30		0	0.00
7	0.35		0	0.00
8	0.40		0	0.00
9	0.45		1	0.01
10	0.50		4	0.04
11	0.55		9	0.09
12	0.60		13	0.13
13	0.65		18	0.18
14	0.70		23	0.23
15	0.75		17	0.17
16	0.80		8	0.08
17	0.85		4	0.04
18	0.90		3	0.03
19	0.95		0	0.00
20	1.00		0	0.00
Total			100	1.00

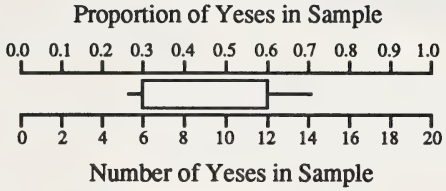
For solutions to **Activity 2**, turn to **Appendix A, Topic 2**.

Activity 3

Use a 90% box and whisker plot chart to find the confidence interval for a survey result.

In the last activity you constructed a box and whisker plot that contained 90% (or more) of the pieces of samples in the box. In this activity you are going to learn how to read charts of 90% box and whisker plots.

Start by taking a closer look at a box and whisker plot that contains 90% of the values in the box.

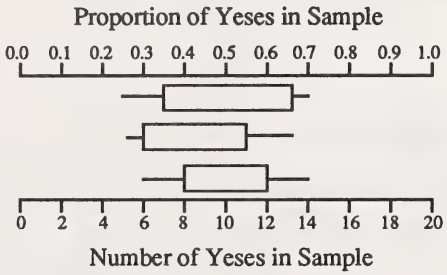


This 90% box and whisker plot is constructed to contain the middle 90% of the sample proportions in the box. The sample proportions that fall within the box will be called **likely sample proportions.** (The edges of the box will be considered as being inside the box.) The sample proportions that fall within the whiskers will be called **unlikely sample proportions.**

If you were to take further samples, it is possible that you could get a sample proportion that would fall outside of the whiskers. Sample proportions that fall outside of the whiskers will also be called **unlikely sample proportions.** Therefore, unlikely sample proportions are sample proportions that fall outside of the box.

In the box and whisker plot shown, the likely sample proportions will be **between** 0.3 and 0.6 inclusive. The unlikely sample proportions will be **less than** 0.3 and **more than** 0.6.

This box and whisker plot was constructed using a population percentage of 40% and forty samples of size 20. The following box and whisker plots were also constructed using the same dimensions.



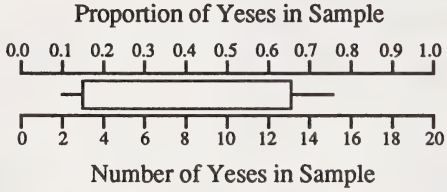
What do these cases tell you?

Each time that you perform the experiment, you get different data; therefore, you get different box and whisker plots.

How can you get the box and whisker plots to be the same?

The key here is in the number of samples. In these cases there were forty samples of size 20 selected. As the number of samples is increased for each of these experiments, the more these box and whisker plots will become similar.

The following is the box and whisker plot for the same experiment, but this time 1000 samples were used.

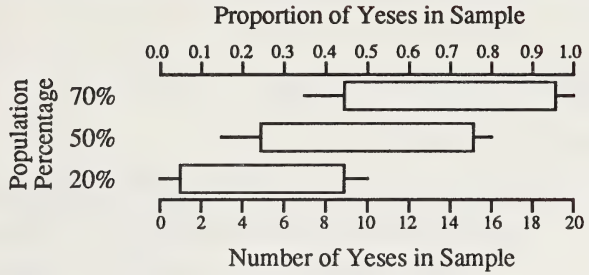


As each of the previous experiments **increases** its number of samples (but keeps the same sample size), the closer its box and whisker plot will come to this box and whisker plot. The **more samples** of a particular size, the **more accurate** the box and whisker plot will be.

What effect does the population percentage have on the box and whisker plot?

Forty samples of size 20 can be explained as follows. Forty samples means the experiment was repeated forty times, while size 20 refers to the number of selections. An example might be selecting twenty marbles from a container that contains a large number of marbles having only two different colours. The twenty marbles are replaced in the container after the results of each draw of twenty are recorded. This is done a total of forty times.

Examine the following box and whisker plots.

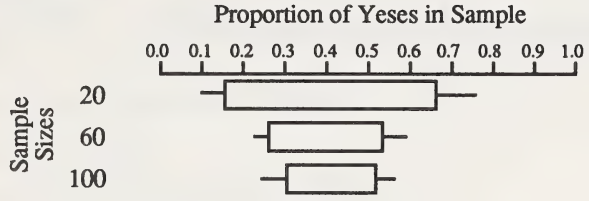


All three of these box and whisker plots were created using 1000 samples of size 20.

As the population percentage **increases**, the box and whisker plot **moves** to the right. Therefore, for every different population percentage, there is a different box and whisker plot.

What effect does the size of the sample have on the box and whisker plot?

Examine the following box and whisker plots that have a population percentage of 40% and are using 1000 samples of different sizes.



The number of yeses in the sample are not given here since each sample size has a different number of yeses for each proportion.

As the sample size gets **larger**, the box becomes **shorter**. Therefore, for every different sample size, there is a different box and whisker plot.

In summary, every different population percentage and every different sample size will require a different box and whisker plot.

A number of these box and whisker plots have already been constructed for you and appear in **Appendix B**. Take a look at them. Notice that there are five different charts. Each chart represents one of the following sample sizes: 20, 40, 60, 80, or 100. Within each chart there are twenty-one different population percentages represented from 0% to 100%.

Take some time to see how you can use these charts.

About 35% of the mathematicians in Northland are women. If a random sample of twenty mathematicians is taken, will the following results be likely or unlikely?

- twelve men and eight women
- four men and sixteen women
- eighteen men and two women

Step 1: Identify the chart that you will be using.

The five charts in **Appendix B** are distinguished by the sample size. In this particular example, a sample of size 20 is being selected since a sample of twenty mathematicians is taken. It is shown on the next page.

Step 2: Select the appropriate parts of the chart.

You are given two different types of information in the question.

- a population percentage
- three different numbers that represent the number of yeses in the sample

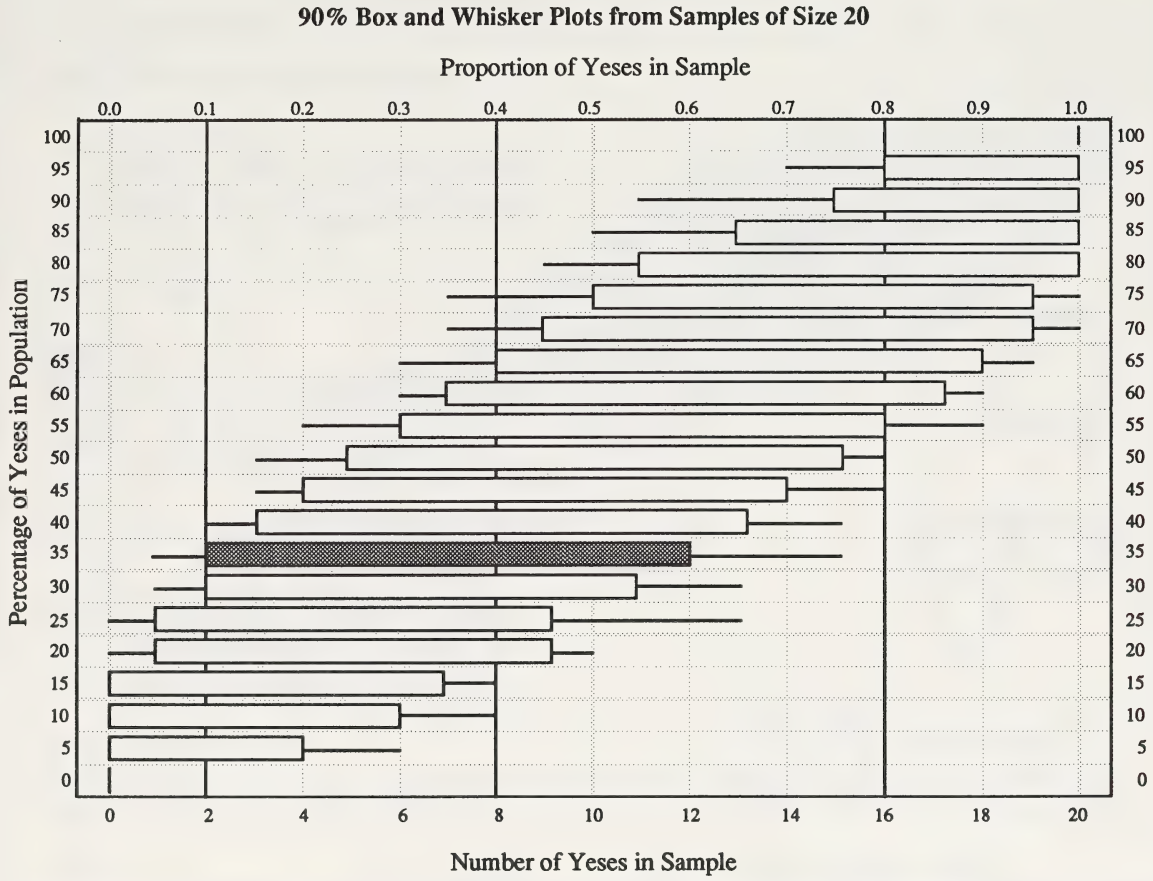
First find 35% on the left or right side of the chart. Move horizontally across the chart to the 90% box and whisker plot that represents a population percentage of 35%. This is the only box and whisker plot that you will be using for this question.

Next find the three different numbers on the chart that represent the number of yeses in the sample (the number of women in the three questions). These three numbers will be along the bottom of the chart. Since this question deals with women, ignore the numbers that deal with men and use only the numbers that apply to women.

Step 3: Decide if the different samples are likely or unlikely.

Move straight up from each of these three points. If the path of this movement goes through the box of the 35% box and whisker plot, the sample is likely. If the path does not go through the box, the sample is unlikely.

This information is summarized on the following chart. Notice that the 35% box and whisker plot has been shaded. Lines have been used to identify the paths from the number of yeses. These paths represent the number of women for each question.



The first path up from eight yeses goes through the box. Therefore, a sample with eight women is likely.

The second path up from sixteen yeses does not go through the box. Therefore, a sample with sixteen women is unlikely.

The third path up from two yeses goes through the left edge of the box. Since the edges of the box are considered as part of the box, a sample with two women is likely.

Take a look at a different type of situation.

A pharmaceutical company has data from a sample of twenty people in which fourteen of the people responded favourably. What population percentages are likely for this situation?

Step 1: Identify the chart that you will be using.

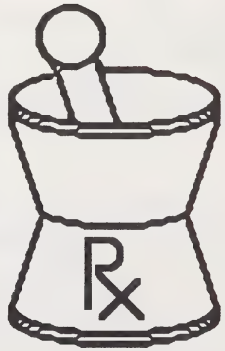
The five charts in **Appendix B** are distinguished by the sample size. In this particular example, a sample of size 20 is being selected.

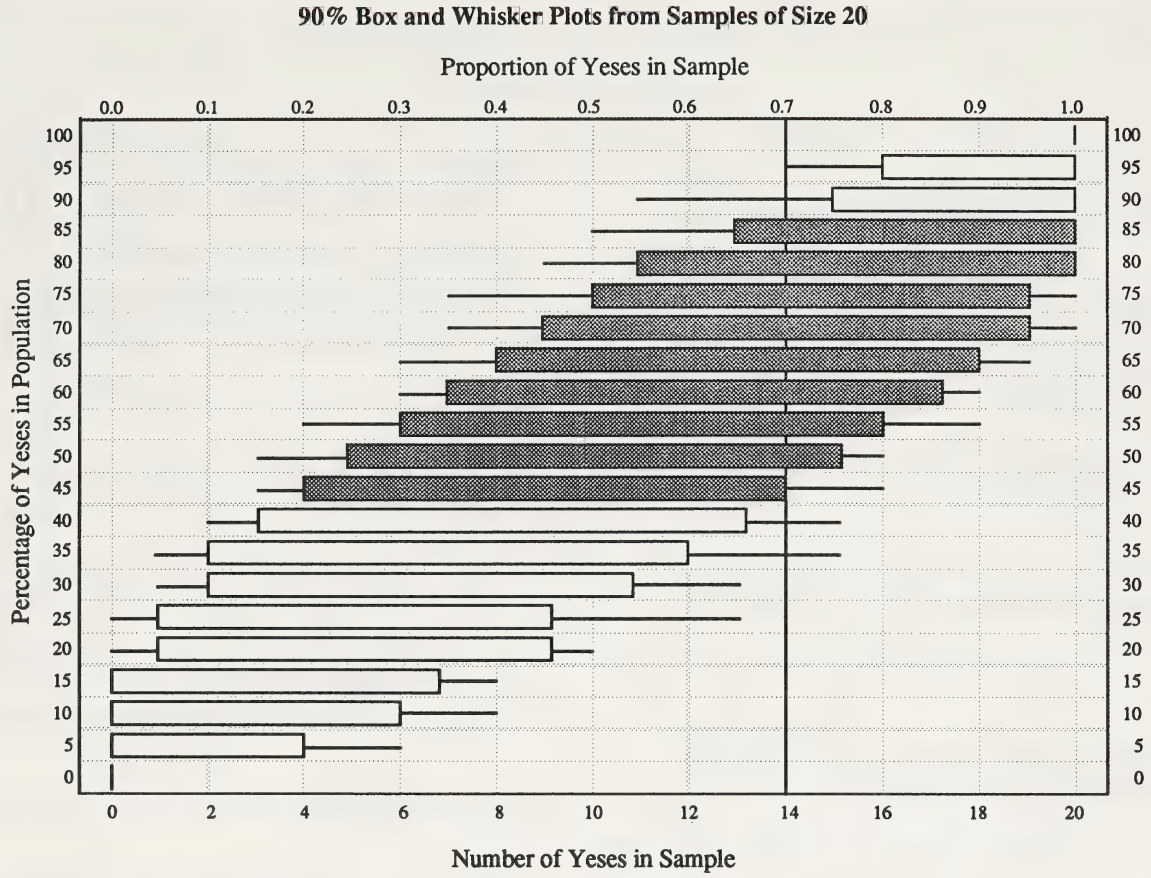
Step 2: Select the appropriate parts of the chart.

In this particular question, the only part that needs to be identified is the number of yeses, which is fourteen.

Step 3: Move straight up and identify the population percentages that are likely for this number of yeses.

Move straight up from fourteen yeses. All of the boxes that the path passes through will be likely population percentages.




90% Box and Whisker Plots from Samples of Size 20
Proportion of Yeses in Sample
0.0
0.1
0.2
0.3
0.4
0.5
0.6
0.7
0.8
0.9
1.0
Percentage of Yeses in Population
100
95
90
85
80
75
70
65
60
55
50
45
40
35
30
25
20
15
10
5
0
0
2
4
6
8
10
12
14
16
18
20
Number of Yeses in Sample

The likely population percentages are 45% to 85% inclusive.

This question has demonstrated the central idea of this topic. The range of population percentages is called the **confidence interval.**

The 90% confidence interval for the pharmaceutical company is 45% to 85%.

What does this mean? If the pharmaceutical company surveyed a sample of twenty people, it would be 90% certain that 45% to 85% of the people would say yes.

The procedure for finding the confidence interval is exactly the procedure that was performed in the last example.

The following are the steps that you will use for finding the confidence interval.

- Identify the appropriate number of yeses or sample proportion.
- Draw a straight vertical line from this point.
- Any box that the line passes through will be part of the confidence interval. Remember that the edges of the boxes are part of the boxes. The whiskers are not used.
- In any concluding statement, make sure to state that out of every 100 trials, 90 trials will succeed and 10 trials will fail.

Try a third and different situation.

What are the likely sample proportions if a sample size of 60 is drawn from a population of 65% yeses?

Step 1: Identify the chart that you will be using.

The five charts in **Appendix B** are distinguished by the sample size. In this particular example, a sample of size 60 is being selected.

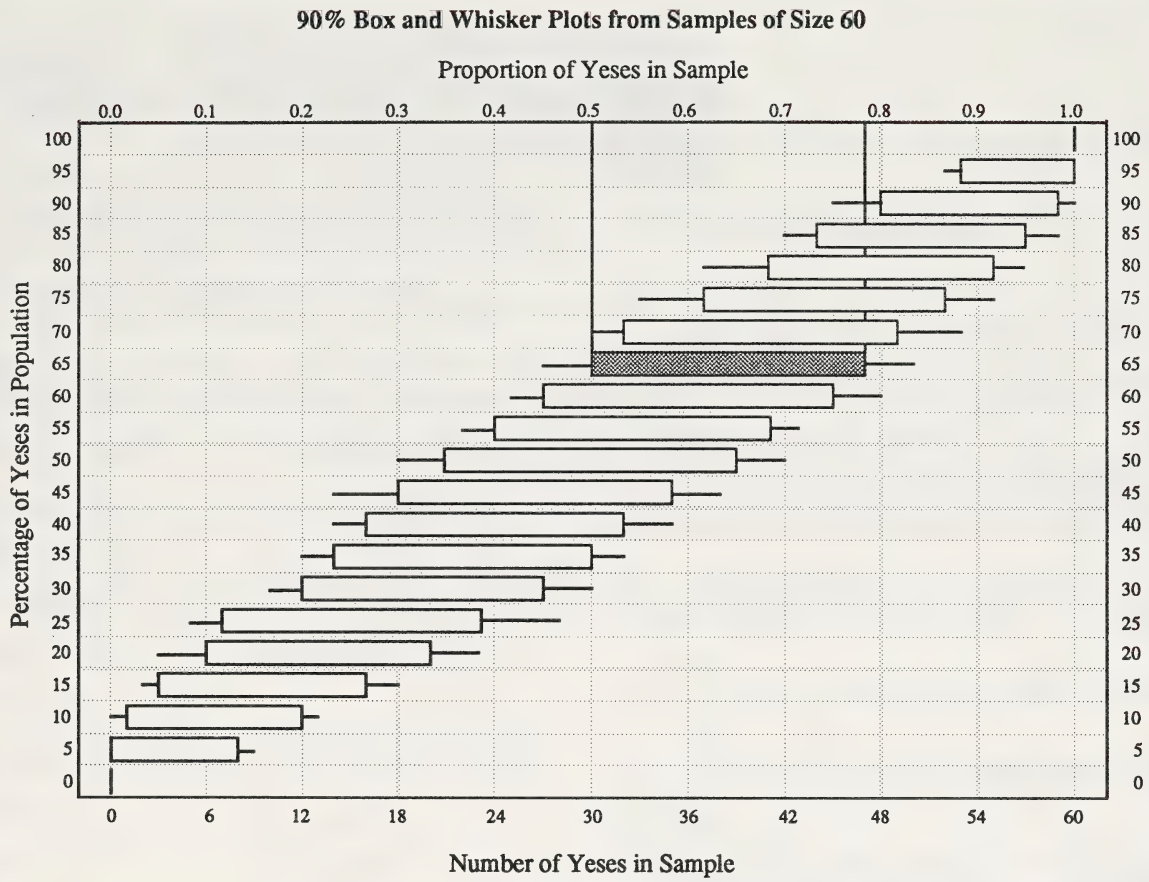
Step 2: Select the appropriate parts of the chart.

In this particular question, the only part that needs to be identified is the population percentage box and whisker plot for 65%.

Step 3: Move straight up from both the left and right of the box and identify the appropriate sample proportions.

This movement is shown in the following chart. The 65% box and whisker plot has been highlighted. Since you are only interested in the sample proportions, it is not necessary to move down to the number of yeses scale. Therefore, the lines only move up to the top scale.

The sample proportions identified in the chart are from 0.5 to 0.78.


90% Box and Whisker Plots from Samples of Size 60
Proportion of Yeses in Sample
0.0
0.1
0.2
0.3
0.4
0.5
0.6
0.7
0.8
0.9
1.0
Percentage of Yeses in Population
100
95
90
85
80
75
70
65
60
55
50
45
40
35
30
25
20
15
10
5
0
0
6
12
18
24
30
36
42
48
54
60
Number of Yeses in Sample

Do any seven of the following questions.

1. A sample size of 40 is selected from a population of 90% yeses. Are each of the following sample proportions likely or unlikely?

   a. 0.25 yeses
   b. 0.65 yeses
   c. 0.825 yeses
   d. 1.00 yeses

2. A random sample of size 80 contains a sample proportion of 0.35 yeses. For which of the following population percentages is this a likely sample proportion?

   a. one with 15% yeses
   b. one with 20% yeses
   c. one with 25% yeses
   d. one with 30% yeses
   e. one with 35% yeses
   f. one with 40% yeses
   g. one with 45% yeses
   h. one with 50% yeses
   i. one with 55% yeses
   j. one with 60% yeses

3. A random sample of size 60 contains seventeen yeses. For which population percentages is this a likely sample proportion?

4. What are the likely sample proportions if you draw a random sample of size 100 from a population with 75% yeses?

5. A recent survey showed that 65% of all motor vehicle accidents on the Great Plains involved alcohol. If a random sample of sixty drivers involved in accidents were taken, determine the following:

   a. Is it likely that a sample proportion of 0.45 will have alcohol in their blood?
   b. Is it likely that forty-eight drivers will have alcohol in their blood?

6. Seventy percent of the people who live in Centreland believe that there should be a national sales tax. If you take a random sample of twenty people from Centreland, determine the following:

   a. Is it likely or unlikely that a sample proportion of 0.95 will approve of the tax?
   b. Is it likely or unlikely that ten people will approve of the tax?
   c. Is it likely or unlikely that a sample proportion of 0.35 will approve of the tax?

7. In a survey, 45% of the people said they would buy a certain new car model constructed by the Northland Engineering Corporation. If a sample size of 40 is used, what percentage of the population will buy the car?

8. What are the confidence intervals for the following situations?

   a. a sample size of 20 and a sample proportion of 0.55
   b. a sample size of 100 and 83 yeses
   c. 70 yeses for a sample size of 80

For solutions to **Activity 3**, turn to **Appendix A, Topic 2.**

Activity 4

Draw statistical conclusions and make inferences to populations, and explain the confidence with which such conclusions and inferences are made based on the results of yes/no surveys.

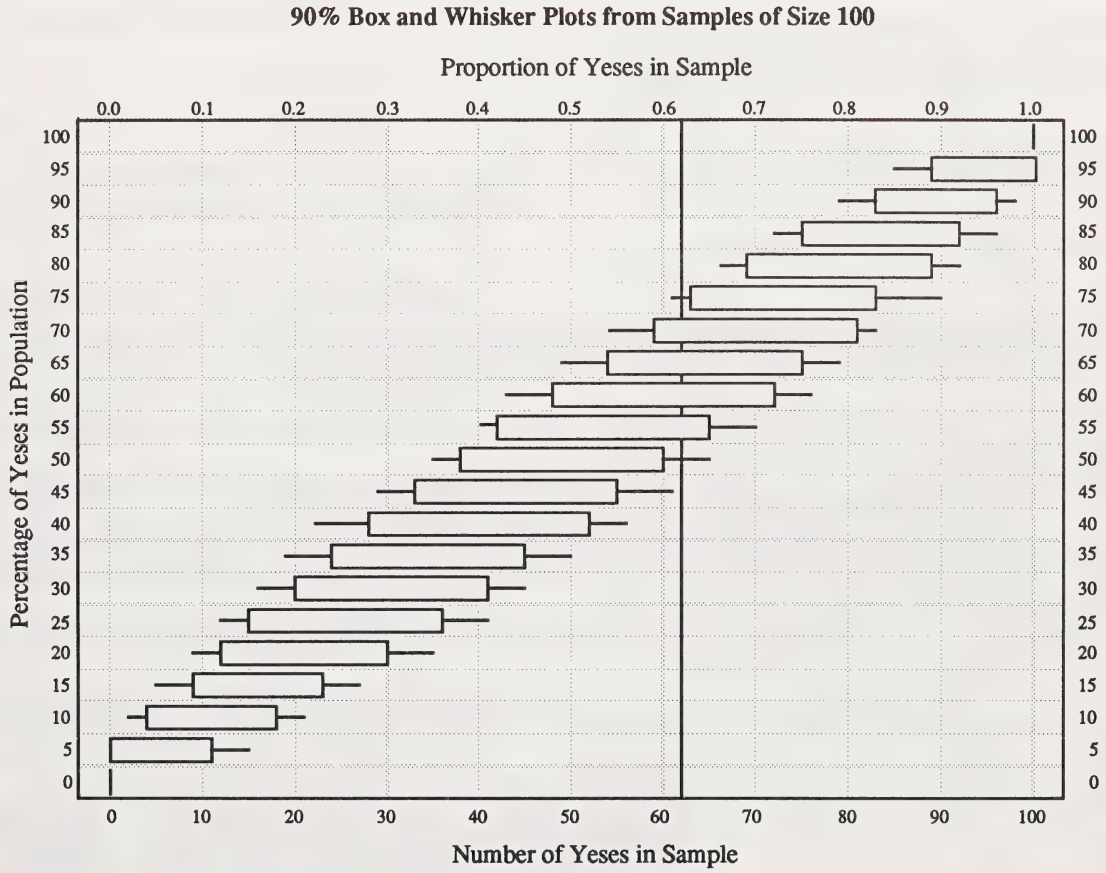
In this activity you are going to employ the tools that you have learned to use in the last three activities to draw conclusions and inferences about real-world situations.

Go through an example and see how these tools are used to draw a proper conclusion.

Student Power magazine gave 100 students two different erasable pens to use for a week: Tic Trick and Matey Peel. The students used the two pens for a week while doing their usual schoolwork. At the end of the week, Student Power surveyed the students to discover which pen they preferred. According to the results, sixty-two students preferred the Tic Trick over the Matey Peel pen.

Assuming that the students were selected randomly, what conclusions would you draw?

First examine the box and whisker chart for a sample size of 100. Drawing a line through 62 yeses, you find the confidence interval. For this particular case the line goes through the boxes for the population percentages 55% to 70%. This chart is as follows.


90% Box and Whisker Plots from Samples of Size 100
Proportion of Yeses in Sample
0.0
0.1
0.2
0.3
0.4
0.5
0.6
0.7
0.8
0.9
1.0
Percentage of Yeses in Population
100
95
90
85
80
75
70
65
60
55
50
45
40
35
30
25
20
15
10
5
0
0
10
20
30
40
50
60
70
80
90
100
Number of Yeses in Sample

Then the following conclusion is drawn.

> If all of the 100 students of a random sample were asked whether they preferred the Tic Trick pen or the Matey Peel pen, 55% to 70% of the students would select the Tic Trick 90% of the time. If 100 groups of 100 students were asked, 90 groups would have 55% to 70% of the students choosing the Tic Trick over the Matey Peel.

The concluding statement to this situation and all concluding statements must have the following information.

- the size of the sample
- the confidence interval
- how many times this confidence interval will happen

In this topic you have been working with 90% confidence intervals. This means that your results will be wrong ten times out of every 100 trials. For the 95% confidence intervals, the results will be wrong only five times out of every 100 trials.

The concluding statements that you are most likely to see in various statistics will deal with 95% confidence intervals. These confidence intervals are calculated using the same procedure that you used in calculating 90% confidence intervals. However, in the exercises for this course you will use only 90% confidence intervals.

Try any five of the following questions.

1. If you take a random sample of size 20 from a population with 35% yeses, is 0.70 a likely sample proportion?

2. In order to be 90% confident, how many samples out of 500 will be likely?

3. If you want to increase the length of the confidence interval, should you increase or decrease the sample size?

4. Fast Food Inc. surveyed a group of sixty people. The company asked the people if they ate breakfast every morning. Write the concluding statement for the survey if forty-nine of the people responded yes to the question.

5. Runners surveyed a group of forty people. The people were asked if they favoured running the public schools year-round. Write the concluding statement for the survey if seventeen of the people responded yes to the question.

6. A group of eighty people was surveyed to determine whether they wanted more parks. Write the concluding statement for the survey if seventy-three of the people responded yes to the question.

For solutions to **Activity 4**, turn to **Appendix A, Topic 2.**

Activity 5

Assess the strengths, weaknesses, and biases of given samples.

In this activity you are going to be looking at the different ways errors can be made in surveys.

Most errors are made during the sampling process. You will examine different types of sampling and the procedures that should be followed when you are conducting a survey.

Sampling

Throughout this topic it has been important that the samples were collected randomly. If a **random sample** was not selected, the method of constructing confidence intervals would not be legitimate.

What is a **random sample**? For a sample to be selected randomly, the following two conditions must be met.

- Each member of the population must have the same chance or probability of being selected.
- Each member of the sample must be selected independently of the others.

It is important to note that these conditions control how the sample is selected, not the results of the selection.

Look at an example.

Twenty students are going to be selected from a school of 300 students.

The name of each student is placed on a 5 cm by 10 cm card. The cards are placed in a large box and shaken. A blindfolded person selects twenty cards from the box, one card at a time.

Is this a random selection? This question can be answered by checking to see if the two conditions have been met.

- Since all of the cards were the same size and the selector had no way of distinguishing among the cards, every student had the same chance of being selected.
- Since only one card was selected each time, there was no way the selection of one card influenced the selection of another card. Therefore, each selection was made independently of the other selections.

Therefore, this is a random selection.

Now suppose that all of the students selected were members of the swimming club. Were the students still selected randomly?

Of course they were. The conditions of random selection were met as shown. The results of the selection do not determine whether a random selection was made.

Since the students are being selected randomly, you would expect the sample to be representative of the population. That is, if all of the twenty students were members of the swimming club, then you would expect all of the students in the population to be members of the swimming club.

Making a random selection is not as easy as it sounds. Suppose you want to take a random sample of the citizens of the city of Red Deer. To make sure that you get a random sample, you will need the names of everyone who lives in Red Deer.

Compiling this material will be very time-consuming. You will have to go to every household in the city and find the names of everyone who lives there. There will be times when no one is home. There will be times when people will give you incorrect information or no information. To make this even worse, people will be moving between places in the city, which may cause you to miss some people or put them on the list twice. In addition, people may be moving in and out of the city, causing some people to be left off the list and other people to be on the list who no longer live in the city.

All of these problems will prevent you from getting a true random sample.

Any list that you find for the people in the city will also have these same problems and others.

Voting lists will list only those people who were eligible to vote at the time of the last election.

City household lists will list only the households; they will not state how many people are in each family or their names.

Telephone directories will list only those individuals who have phones and want their names listed in the directory.

When a method of selection tends to overrepresent or underrepresent some part of the population, the method is said to **bias** the sample. As shown, it is almost impossible to make a selection without some bias. If it is impossible for you to make a selection without a bias, then make the selection as randomly as possible.

Since it is not possible to get a truly random sample of the entire population, different types of sampling procedures are used to gather a sample. Now you will look at some of these sampling techniques and see how they can be used to create a biased sample.

Convenience Sampling: Each member of the sample is simply chosen without the use of any random mechanism.

A convenience sample could be selected by picking the first twenty people you see, or the first twenty people who talk to you, or the twenty tallest people, etc.

Since convenience sampling uses no explicit random mechanism to select the sample, no confidence interval statements can be made about the population.

Convenience samples are usually biased.

Self-Selected Sampling: Each member of the sample has volunteered to be in the survey. Self-selected sampling is a type of convenience sampling.

The most common type of self-selected sampling is returning a survey form from a newspaper or magazine.

People who volunteer to participate in a survey usually are the most concerned and motivated people in the population.

Judgement Sampling: Each member of the sample has been selected by an expert or someone who is knowledgeable about the population. Judgement sampling is a type of convenience sampling.

Judgement sampling is done by a produce purchaser. The purchaser will select and taste several fruits from a shipment in order to decide on the quality of the shipment. This type of sampling sometimes can give you a better measurement of the quality of a population than a random sample.

Probability Sampling: Each member of the sample is selected by use of a probability mechanism. Also, each member of the population may not have the same chance of being selected or of being selected independently.

Probability sampling may or may not be a random sample. This decision is made with the selection of the probability mechanism.

Statisticians have developed methods for calculating confidence intervals from probability samples, but these methods are complicated and will not be discussed in this course.

Remember: Your work with the confidence interval depended on the sample being selected randomly.

Clustered Sampling: A group is selected randomly from the population and then every member of the group becomes a member of the sample. This is a type of probability sampling, but it is not simple random sampling.

A town may use a clustered sample by randomly selecting streets, then randomly selecting households along those streets, and finally surveying everyone who lives in those households.

Clustered sampling is not a type of convenience sampling since at no time does the surveyor decide who will be included in the sample. Since clusters or groups are being selected, not all of the members of the population have the same chance of being selected. Also, since the group is bound together by some common trait, the members of the sample are not being selected independently.

Stratified Random Sampling: Each member of the population is placed into a stratum or subgroup, and then the members of the sample are found by taking a random selection from each stratum or subgroup. This is a type of probability sampling, but it is not simple random sampling.

An example of stratified random sampling occurs when a school is divided into a group of boys and a group of girls, and then twenty people are selected from each of the two groups. Another example occurs when the school is divided into its grades, and then twenty students are selected from each grade.

A stratified sample may be used when you want to know how a certain stratum replies in comparison to the whole population. As an example, you may want to know not only how the whole student body answers a survey, but how both the boys and the girls answer that survey.

A stratified sample also may be used when you want to ensure that certain strata have their opinions heard. This will cause the sample to be more representative of the entire population. The increased representation causes the confidence interval for the stratified sample to be shorter than the confidence interval from a random sample of the same population. This will cause the stratified sample to give more precise estimates than a random sample.

Systematic Sampling: Each member of the sample is selected from the population by using a system. This is a type of probability sampling.

One type of systematic sampling is to number each member of the population, and then select each *n*th (fourth, eighth, or tenth, for example) member of the population to be part of the sample. An example would be to select every fifth student who walks into your classroom.

Systematic sampling has several advantages over random sampling. It is easier to do than random sampling and it guarantees that the sample is taken from throughout the population. There are disadvantages to systematic sampling. It is possible that certain characteristics may be able to hide in the population using this method. An example of this would be the case where a street intersection is being studied. If the intersection is only studied on Monday and Tuesday, the study will not cover the traffic problems that show up on Friday and Saturday.

Systematic sampling is most useful when random sampling is too difficult to be carried out and the ordering of the population does not create a problem.

Other Forms of Bias

Even if an organization uses random sampling or probability sampling to choose a sample, the survey can become biased for other reasons.

- People may refuse or neglect to respond.
  - The question might be sensitive to some people.
  - The interviewer may not be able to contact some people.

- People may be untruthful.
  - People may be trying to give socially acceptable responses.
  - People might be trying to tell the interviewer the answers that they believe the interviewer wants to hear.
  - People might be trying to hide the fact that they do not know the answer to the question.
  - People may inadvertently improve or change a numerical response.

- The survey may be conducted at a poor time.
  - There may have just been a campaign concerning the topic of the survey.

- The survey may be poorly worded.
  - People are more willing to accept certain phrases. For example, people are more willing to accept the phrase "not allow" than "forbid."
  - The surveyor may word the question to cause people to respond in a certain way.

- The interviewers may not be properly trained.
  - An interviewer can easily misinterpret people's responses.

- Errors may be made in repetitive tasks such as computer entry errors or simple tabulation errors.

The following are some hints for making your survey as unbiased as possible.

- Make sure that a good cross section of the population is selected for the survey.

- Do a follow-up to the survey to make sure that all the people in the sample have responded.

- Take extreme care in the wording of the questions to eliminate all ambiguities and to avoid using socially sensitive words.

- Ensure anonymity in the survey. Some people will not respond if their responses can be identified.

- Make sure that the question asks for a yes/no response. Many people will not take the time to write out responses.

- In interviews, the interviewer must be careful not to emphasize any one word or phrase. If the respondent fails to understand the question, the interviewer must only repeat the question and not explain or interpret the question.

Answer any five of questions 1 to 6, and question 7.

1. Which of the following situations is a random sampling of students from a high school? Explain why any situation is not a random sampling.

   a. Select the students whose last digit in their phone number is a 6.

   b. Assign each student a number and then use a random number table to select the students.

   c. Select only those students whose name starts with the letters A, B, C, or D.

2. A sample of 200 people is selected from across Canada. Three people live on the same street. Is it possible that this is a random sample? Explain your answer.

3. Explain why you would not be able to obtain a true random sample of the people of Calgary.

4. State what type of sampling is being used in each of the following cases.

   a. Readers are asked to return a survey that was given in a magazine.

   b. One student from each home classroom in a school is selected to be on a social committee.

   c. A buyer for a cannery tries several different pieces of fruit to judge its quality.

5. Explain three different situations in which a person might give an untruthful answer to a survey.

6. Explain why a survey might report that 5% more women than men responded that they were married.

7. Plan and carry out your own survey.

   a. Write a yes/no-type survey question on a topic that you find interesting.

   b. Decide what your population will be. As an example, you may use your school or your community as your population.

   c. Try your survey question on a few people that you know to make sure that they are interpreting the question as you planned.

   d. Decide on a method and select a random sample of forty people. Explain your method.

   e. Survey the forty people in your sample. Tabulate the results.

   f. What is the sample proportion for your survey?

   g. What is the 90% confidence interval for the sample proportion of your survey?

   h. Write a concluding statement to explain the findings of your survey.

For solutions to **Activity 5**, turn to **Appendix A, Topic 2**.

If you require help, do the Extra Help section.

If you want more challenging explorations, do the Extensions section. } You may decide to do both.

Extra Help

Constructing a 90% Box and Whisker Plot Chart

In Activity 2 you were given the following steps for creating a 90% box and whisker plot.

Step 1: Construct two axes. The first one is for the proportion of yeses in the sample, and the second one is for the number of yeses in the sample.

Step 2: Find the number of values that will be in the box and the whiskers. Since you are making a 90% box and whisker plot, 90% of the values will be in the box and 5% of the values will be in each whisker on both sides of the box. The box can have more than 90% of the values, but any one whisker will have at most 5% of the values.

Step 3: Identify the number of yeses that will mark the ends of the box and the whiskers.

Step 4: Find the ends of the box and the whiskers on the axes. Mark these locations with dots.

Step 5: Using the dots as your guides, draw in the box and the whiskers.

The following is part of the sampling distribution for a population percentage of 5% yeses. The five steps shown will be used to make the 90% box and whisker plot for this information.

Number of Yeses	Sample Proportion	Frequency	Proportion of All Trials
0	0.00	355	0.355
1	0.05	320	0.320
2	0.10	179	0.179
3	0.15	92	0.092
4	0.20	52	0.052
5	0.25	1	0.001
6	0.30	1	0.001
7	0.35	0	0.000
8	0.40	0	0.000
9	0.45	0	0.000
.	.	.	.
.	.	.	.
.	.	.	.
20	1.00	0	0.000
Total		1000	1.000

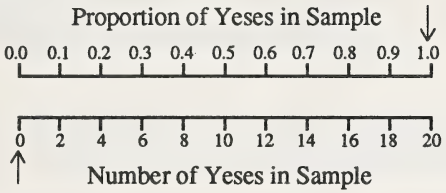
Step 1

Construct two axes. The first one is for the proportion of yeses in the sample, and the second one is for the number of yeses in the sample.

The axis for the proportion of yeses will represent what percentage of the sample is made up of yeses, where 0 will represent the case where there are no yeses in the sample, and 1 will represent the case where the entire sample is made of yeses.

The second axis will represent the number of yeses in the sample.

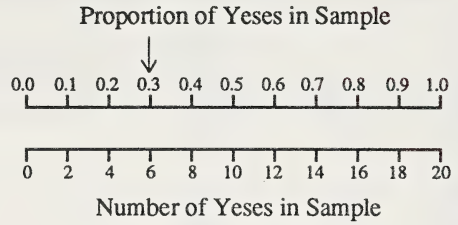
Since there are twenty responses in the sample, twenty yeses will correspond to the number 1 on the proportion of yeses scale. Therefore, you must make sure that these two numbers line up on the chart.



Also note that the zeros on both axes line up.

Choose any number along the bottom scale. Divide it by 20. Find this number along the top scale. Go straight down to the bottom scale. Is this the number that you chose? It should be.

For example, start with the number 6. Dividing 6 by 20 will give you 0.3. Find 0.3 along the top axis. From here move straight down to the bottom axis. Are you back at the number 6? Check the following axes.



The axes are now ready.

Step 2

Find the number of values that will be in the box and the whiskers. Since you are making a 90% box and whisker plot, 90% of the values will be in the box and 5% of the values will be in each whisker on both sides of the box.

The sampling distribution has 1000 values. If 90% of the values are to be in the box, then 90% of 1000 or 900 values will be in the box.

That leaves 100 values outside of the box, or 50 values on each side of the box.

Step 3

Identify the number of yeses that will mark the ends of the box and the whiskers.

Count down from the top of the frequency column in the sampling distribution until you get to the number 50. The rows you have covered in the sampling distribution will represent the left whisker of the plot. If the row that you are in adds up to 51 or more, that category will be part of the box. Remember that the box must contain at least the middle 900 of the values from the sampling distribution table. The box can contain more than 900 of the values, but it must not contain fewer.

Since the first row contains more than fifty values, this plot will not have a left whisker. The box will start at a point on the far left of the axis.

Number of Yeses	Sample Proportion	Frequency	Proportion of All Trials	
0	0.00	355	0.355	} Since this row already contains more than fifty values, there will be no left whisker.
1	0.05	320	0.320	
2	0.10	179	0.179	
3	0.15	92	0.092	
4	0.20	52	0.052	
5	0.25	1	0.001	
6	0.30	1	0.001	

Now count up from the bottom of the frequency column. Again you want the frequencies to add up to 50, but not more than 50. These rows will represent the right whisker of the plot.

Number of Yeses	Sample Proportion	Frequency	Proportion of All Trials	
0	0.00	355	0.355	
1	0.05	320	0.320	
2	0.10	179	0.179	
3	0.15	92	0.092	
4	0.20	52	0.052	
5	0.25	1	0.001	} These two rows will represent the right whisker. The next row will not be part of the whisker since it will add up to 54, which is more than 50.
6	0.30	1	0.001	

The remaining rows will be part of the box.

Number of Yeses	Sample Proportion	Frequency	Proportion of All Trials	
0	0.00	355	0.355	} These rows will be represented by the box in the chart.
1	0.05	320	0.320	
2	0.10	179	0.179	
3	0.15	92	0.092	
4	0.20	52	0.052	
5	0.25	1	0.001	
6	0.30	1	0.001	

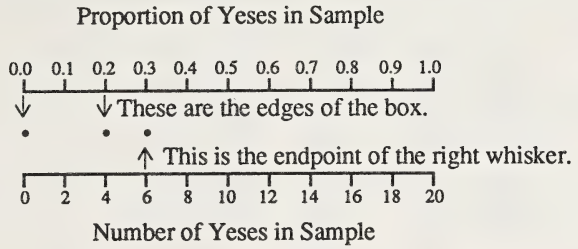
From this you can see that the left edge of the box will be at 0 yeses, the right side of the box will be at 4 yeses, and the right edge of the right whisker will be at 6 yeses.

Now you are ready to move on to the fourth step.

Step 4

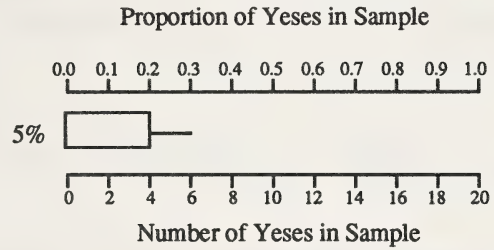
Find the ends of the box and the whiskers on the axes. Mark these locations with dots.

This particular set of axes will have only three dots. The first two dots will mark the locations of the edges of the box and the third dot will mark the endpoint of the right whisker.



Step 5

Using the dots as your guides, draw in the box and the whiskers.



The 90% box and whisker plot for the 5% population percentage is now complete.

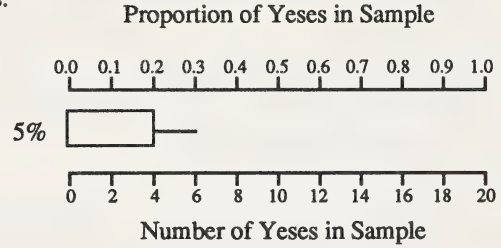
The following sampling distribution is used to find the 90% box and whisker plot for a population percentage of 10%. Use this information and draw the 90% box and whisker plot.

Number of Yeses	Sample Proportion	Frequency	Proportion of All Trials
0	0.00	136	0.136
1	0.05	264	0.264
2	0.10	269	0.269
3	0.15	195	0.195
4	0.20	96	0.096
5	0.25	20	0.020
6	0.30	12	0.012
7	0.35	6	0.006
8	0.40	2	0.002
9	0.45	0	0.000
.	.	.	.
.	.	.	.
.	.	.	.
20	1.00	0	0.000
Total		1000	1.000

Step 1

Construct two axes. The first one is for the proportion of yeses in the sample, and the second one is for the number of yeses in the sample.

You will graph the 10% yeses plot with the 5% yeses plot using the same axes.



Step 2

Find the number of values that will be in the box and the whiskers. Since you are making a 90% box and whisker plot, 90% of the values will be in the box and 5% of the values will be in each whisker on both sides of the box.

This box and whisker plot also will have at least 900 values in the box and no more than 50 values for each whisker.

Step 3

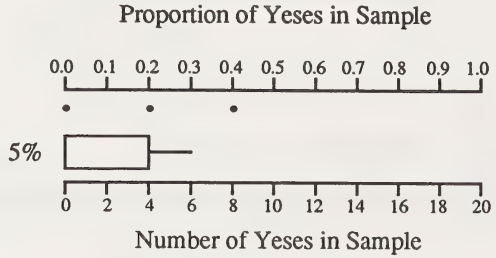
Identify the number of yeses that will mark the ends of the box and the whiskers.

Number of Yeses	Sample Proportion	Frequency	Proportion of All Trials	
0	0.00	136	0.136	These values represent the box.
1	0.05	264	0.264	
2	0.10	269	0.269	
3	0.15	195	0.195	
4	0.20	96	0.096	
5	0.25	20	0.020	These values represent the right whisker.
6	0.30	12	0.012	
7	0.35	6	0.006	
8	0.40	2	0.002	
9	0.45	0	0.000	

The ends of the box will be at 0 yeses and 4 yeses, and the right end of the right whisker will be at 8 yeses.

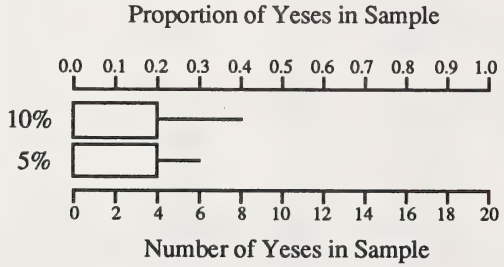
Step 4

Find the ends of the box and the whiskers on the axes. Mark these locations with dots.



Step 5

Using the dots as your guides, draw in the box and the whiskers.



This box and whisker plots for the 5% and 10% population percentages are now complete. This process could be continued for every 5% up to and including 95%.

Draw box and whisker plots for any four of the following 90% box and whisker plots.

1. The following is for a population percentage of 15%. The sample size is 20.

Number of Yeses	Sample Proportion	Frequency	Proportion of All Trials
0	0.00	55	0.055
1	0.05	149	0.149
2	0.10	166	0.166
3	0.15	215	0.215
4	0.20	169	0.169
5	0.25	132	0.132
6	0.30	59	0.059
7	0.35	47	0.047
8	0.40	8	0.008
9	0.45	0	0.000
.	.	.	.
.	.	.	.
.	.	.	.
20	1.00	0	0.000
Total		1000	1.000

2. The following is for a population percentage of 20%. The sample size is 20.

Number of Yeses	Sample Proportion	Frequency	Proportion of All Trials
0	0.00	17	0.017
1	0.05	59	0.059
2	0.10	136	0.136
3	0.15	183	0.183
4	0.20	170	0.170
5	0.25	141	0.141
6	0.30	103	0.103
7	0.35	77	0.077
8	0.40	63	0.063
9	0.45	49	0.049
10	0.50	2	0.002
11	0.55	0	0.000
.	.	.	.
.	.	.	.
.	.	.	.
20	1.00	0	0.000
Total		1000	1.000

3. The following is for a population percentage of 25%. The sample size is 20.

Number of Yeses	Sample Proportion	Frequency	Proportion of All Trials
0	0.00	4	0.004
1	0.05	47	0.047
2	0.10	82	0.082
3	0.15	117	0.117
4	0.20	151	0.151
5	0.25	157	0.157
6	0.30	138	0.138
7	0.35	117	0.117
8	0.40	101	0.101
9	0.45	49	0.049
10	0.50	20	0.020
11	0.55	11	0.011
12	0.60	5	0.005
13	0.65	1	0.001
14	0.70	0	0.000
.	.	.	.
.	.	.	.
.	.	.	.
20	1.00	0	0.000
Total		1000	1.000

4. The following is for a population percentage of 30%. The sample size is 20.

Number of Yeses	Sample Proportion	Frequency	Proportion of All Trials
0	0.00	0	0.000
1	0.05	22	0.022
2	0.10	39	0.039
3	0.15	54	0.054
4	0.20	93	0.093
5	0.25	107	0.107
6	0.30	123	0.123
7	0.35	140	0.140
8	0.40	134	0.134
9	0.45	108	0.108
10	0.50	83	0.083
11	0.55	57	0.057
12	0.60	32	0.032
13	0.65	8	0.008
14	0.70	0	0.000
.	.	.	.
.	.	.	.
.	.	.	.
20	1.00	0	0.000
Total		1000	1.000

5. The following is for a population percentage of 35%. The sample size is 20.

Number of Yeses	Sample Proportion	Frequency	Proportion of All Trials
0	0.00	0	0.000
1	0.05	7	0.007
2	0.10	52	0.052
3	0.15	78	0.078
4	0.20	102	0.102
5	0.25	111	0.111
6	0.30	123	0.123
7	0.35	129	0.129
8	0.40	113	0.113
9	0.45	98	0.098
10	0.50	76	0.076
11	0.55	57	0.057
12	0.60	42	0.042
13	0.65	11	0.011
14	0.70	0	0.000
15	0.75	1	0.001
16	0.80	0	0.000
.	.	.	.
.	.	.	.
.	.	.	.
20	1.00	0	0.000
Total		1000	1.000

For solutions to **Extra Help**, turn to **Appendix A, Topic 2**.

Extensions

Using a Microcomputer to Simulate Yes/No Responses

In Activity 1 you were asked to use a microcomputer to simulate the responses to a yes/no question from forty different samples. Each sample was to contain thirty responses and the population percentage was to be 43%.

Look at how this can be done using the BASIC computing language of the Apple II [1] series microcomputer (Applesoft).

- Have the microcomputer select a number between 0 and 99 inclusive.

140 S(A, B) = INT(100*RND (1))

The command RND (1) will select a random number between 0 and 1 with nine digits after the decimal point. This number is multiplied by 100, giving you a number between 0 and 100 with seven digits after the decimal point. The command INT will cut off the digits after the decimal. This will leave you with a whole number between 0 and 99 inclusive. The number will be stored in the location called S(A, B).

[1] Apple II ™ is a trademark of Apple Computers, Inc.

- Have the microcomputer select thirty numbers for each sample.

```
130 FOR B = 1 TO 30
140 S(A, B) = INT(100*RND (1))
180 NEXT B
```

This FOR/NEXT loop will cause the microcomputer to select thirty different numbers and store them in the locations S(A, 1) to S(A, 30).

- Have the microcomputer select forty samples of thirty numbers.

```
120 FOR A = 1 TO 40
130 FOR B = 1 TO 30
140 S(A, B) = INT(100*RND (1))
180 NEXT B
190 NEXT A
```

This FOR/NEXT loop will cause the microcomputer to select forty different samples of thirty numbers and store them in the locations S(1, 1) to S(40, 30).

- Now that the numbers have been selected, have the microcomputer decide what response each number represents, either yes or no.

```
120 FOR A = 1 TO 40
130 FOR B = 1 TO 30
140 S(A, B) = INT(100*RND (1))
150 IF S(A, B) < 43 THEN S$(A, B) = "Y"
160 IF S(A, B) >= 43 THEN S$(A, B) = "N"
180 NEXT B
190 NEXT A
```

In this particular case, if the number is less than 43, then the responses are yes. If the number of responses is greater than or equal to 43, then the responses are no. The responses are stored in the locations S$(1, 1) to S$(40, 30).

- Have the microcomputer count the number of positive responses for each sample.

```
120 FOR A = 1 TO 40
130 FOR B = 1 TO 30
140 S(A, B) = INT(100*RND (1))
150 IF S(A, B) < 43 THEN S$(A, B) = "Y"
160 IF S(A, B) >= 43 THEN S$(A, B) = "N"
170 IF S(A, B) < 43 THEN N(A) = N(A) + 1
180 NEXT B
190 NEXT A
```

Here the microcomputer will increase the number located at N(A) by one each time a positive response is encountered.

- Since the responses are being stored in an array, the microcomputer must be informed of the sizes of the arrays. This allows the microcomputer to set aside the memory to store the responses in these locations.

```
110 DIM N(40), S(40, 30), S$(40, 30)
120 FOR A = 1 TO 40
130 FOR B = 1 TO 30
140 S(A, B) = INT(100*RND (1))
150 IF S(A, B) < 43 THEN S$(A, B) = "Y"
160 IF S(A, B) >= 43 THEN S$(A, B) = "N"
170 IF S(A, B) < 43 THEN N(A) = N(A) + 1
180 NEXT B
190 NEXT A
```

The numbers that are replacing the letters in this line are informing the microcomputer of how much space is going to be required for these arrays.

- Since you do not want any errant data interfering with your work, have the microcomputer clear out all numbers that might be lingering in its memory.

```
100 CLEAR
110 DIM N(40), S(40, 30), S$(40, 30)
120 FOR A = 1 TO 40
130 FOR B = 1 TO 30
140 S(A, B) = INT(100*RND (1))
150 IF S(A, B) < 43 THEN S$(A, B) = "Y"
160 IF S(A, B) >= 43 THEN S$(A, B) = "N"
170 IF S(A, B) < 43 THEN N(A) = N(A) + 1
180 NEXT B
190 NEXT A
```

This command will cause all numbers in all memory locations to be set to zero.

When this program is RUN, the microcomputer will simulate all of the responses. The only remaining problem is that the microcomputer has not been told to tell you the responses.

- Tell the microcomputer to print out a response.

```
100 CLEAR
110 DIM N(40), S(40, 30), S$(40, 30)
120 FOR A = 1 TO 40
130 FOR B = 1 TO 30
140 S(A, B) = INT(100*RND (1))
150 IF S(A, B) < 43 THEN S$(A, B) = "Y"
160 IF S(A, B) >= 43 THEN S$(A, B) = "N"
170 IF S(A, B) < 43 THEN N(A) = N(A) + 1
180 NEXT B
190 NEXT A
230 PRINT S$(A, B);
```

The semicolon at the end of the line will cause the next response to be printed right beside the previous response.

- Tell the microcomputer to print out all thirty of the responses in a sample.

```
100 CLEAR
110 DIM N(40), S(40, 30), S$(40, 30)
120 FOR A = 1 TO 40
130 FOR B = 1 TO 30
140 S(A, B) = INT(100*RND (1))
150 IF S(A, B) < 43 THEN S$(A, B) = "Y"
160 IF S(A, B) >= 43 THEN S$(A, B) = "N"
170 IF S(A, B) < 43 THEN N(A) = N(A) + 1
180 NEXT B
190 NEXT A
220 FOR B = 1 TO 30
230 PRINT S$(A, B);
240 NEXT B
```

- Have the microcomputer print out all forty of the samples.

```
100 CLEAR
110 DIM N(40), S(40, 30), S$(40, 30)
120 FOR A = 1 TO 40
130 FOR B = 1 TO 30
140 S(A, B) = INT(100*RND (1))
150 IF S(A, B) < 43 THEN S$(A, B) = "Y"
160 IF S(A, B) >= 43 THEN S$(A, B) = "N"
170 IF S(A, B) < 43 THEN N(A) = N(A) + 1
180 NEXT B
190 NEXT A
200 FOR A = 1 TO 40
220 FOR B = 1 TO 30
230 PRINT S$(A, B);
240 NEXT B
260 NEXT A
```

- At the end of each sample, have the microcomputer print out the number of positive responses in that sample. This line also will prevent all of the responses from all of the samples from being printed together.

```
100 CLEAR
110 DIM N(40), S(40, 30), S$(40, 30)
120 FOR A = 1 TO 40
130 FOR B = 1 TO 30
140 S(A, B) = INT(100*RND (1))
150 IF S(A, B) < 43 THEN S$(A, B) = "Y"
160 IF S(A, B) >= 43 THEN S$(A, B) = "N"
170 IF S(A, B) < 43 THEN N(A) = N(A) + 1
180 NEXT B
190 NEXT A
200 FOR A = 1 TO 40
220 FOR B = 1 TO 30
230 PRINT S$(A, B);
240 NEXT B
250 PRINT " "; N(A)
260 NEXT A
```

- You can continue to modify the program by numbering each sample. This will give your presentation a cleaner look.

```
100 CLEAR
110 DIM N(40), S(40, 30), S$(40, 30)
120 FOR A = 1 TO 40
130 FOR B = 1 TO 30
140 S(A, B) = INT(100*RND (1))
150 IF S(A, B) < 43 THEN S$(A, B) = "Y"
160 IF S(A, B) >= 43 THEN S$(A, B) = "N"
170 IF S(A, B) < 43 THEN N(A) = N(A) + 1
180 NEXT B
190 NEXT A
200 FOR A = 1 TO 40
210 PRINT A;" ";
220 FOR B = 1 TO 30
230 PRINT S$(A, B);
240 NEXT B
250 PRINT " "; N(A)
260 NEXT A
```

- A printout of the responses can be obtained by using the following PR commands.

```
100 CLEAR
110 DIM N(40), S(40, 30), S$(40, 30)
120 FOR A = 1 TO 40
130 FOR B = 1 TO 30
140 S(A, B) = INT(100*RND (1))
150 IF S(A, B) < 43 THEN S$(A, B) = "Y"
160 IF S(A, B) >= 43 THEN S$(A, B) = "N"
170 IF S(A, B) < 43 THEN N(A) = N(A) + 1
180 NEXT B
190 NEXT A
195 PR#1
200 FOR A = 1 TO 40
210 PRINT A;" ";
220 FOR B = 1 TO 30
230 PRINT S$(A, B);
240 NEXT B
250 PRINT " "; N(A)
260 NEXT A
265 PR#0
```

PR#1 will send the upcoming information to the printer instead of the screen and PR#0 will send the information to the screen instead of the printer.

The program is now complete.

Answer the following questions.

1. Use the program that was created in this extension to create a new set of data. Once you have created the data, compile the information into a sampling distribution.

2. Make changes to the program that was created in this extension so the microcomputer will select sixty samples with twenty responses in each sample.

For solutions to **Extensions**, turn to **Appendix A, Topic 1**.

Unit Summary

What You Have Learned

In this unit the following concepts were discussed:

- sets of bivariate data plotted to produce a scatterplot
- median fit method for finding the line of best fit on a scatterplot
- the equation of the line of best fit and the use of the equation for predictions
- positive and negative correlations between the variables of a bivariate distribution
- construction of 90% box and whisker plots
- the use of a 90% box and whisker plot chart to find the confidence interval for a survey result
- the confidence with which conclusions and inferences are made based on the results of yes/no surveys

You are now ready to

complete the **Unit Assignment.**

Appendices

Appendix A
Solutions

Review

Topic 1 Bivariate Data

Topic 2 Confidence

Appendix B
Charts and Tables

90% Box and Whisker Plots from Samples of Size 20

90% Box and Whisker Plots from Samples of Size 40

90% Box and Whisker Plots from Samples of Size 60

90% Box and Whisker Plots from Samples of Size 80

90% Box and Whisker Plots from Samples of Size 100

Random Number Table

Appendix A
Solutions

Review

1. a. $\bar{x} = \dfrac{a_1 + a_2 + a_3 + \ldots + a_n}{n}$

$\bar{x} = \dfrac{18+34+14}{3}$

$\bar{x} = 22$

b. $\bar{x} = \dfrac{a_1 + a_2 + a_3 + \ldots + a_n}{n}$

$\bar{x} = \dfrac{6+4+7+10+9+9+11+16}{8}$

$\bar{x} = 9$

c. $\bar{x} = \dfrac{a_1 + a_2 + a_3 + \ldots + a_n}{n}$

$\bar{x} = \dfrac{141+107+118+114+125}{5}$

$\bar{x} = 121$

2. First arrange the values in ascending order, and then select the middle value or take the average of the two middle values.

a. 14, 18, 34

median = 18

b. 4, 6, 7, 9, 9, 10, 11, 16

$\text{median} = \dfrac{9+9}{2} = \dfrac{18}{2}$

$= 9$

c. 107, 114, 118, 125, 141

median = 118

3. a. $m = \dfrac{y_1 - y_2}{x_1 - x_2}$

$m = \dfrac{7-5}{4-9}$

$m = \dfrac{2}{-5}$

$m = \dfrac{-2}{5}$

b. $m = \dfrac{y_1 - y_2}{x_1 - x_2}$

$m = \dfrac{22-102}{108-112}$

$m = \dfrac{-80}{-4}$

$m = 20$

4. a. $y-y_1=\left(\frac{y_2-y_1}{x_2-x_1}\right)(x-x_1)$

$y-7=\frac{-2}{5}(x-4)$

$5y-35=-2x+8$

$2x+5y-43=0$

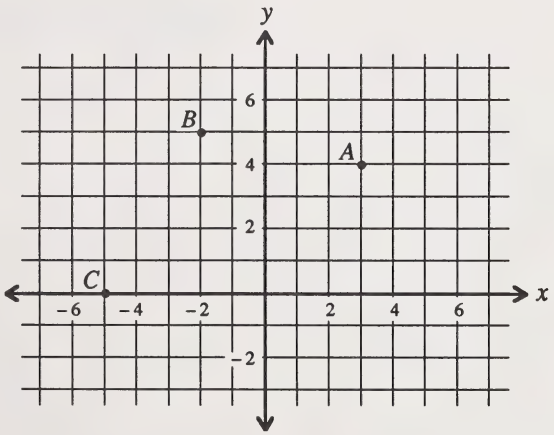
b. $y-y_1=\left(\frac{y_2-y_1}{x_2-x_1}\right)(x-x_1)$

$y-22=20(x-108)$

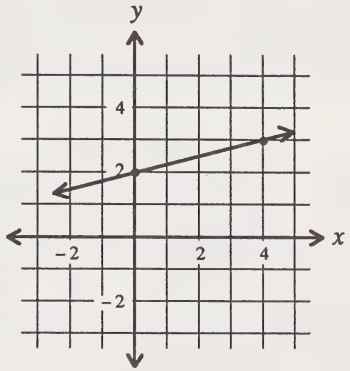
$y-22=20x-2160$

$20x-y-2138=0$

5.



6. a.



The first point is $(0, 2)$.

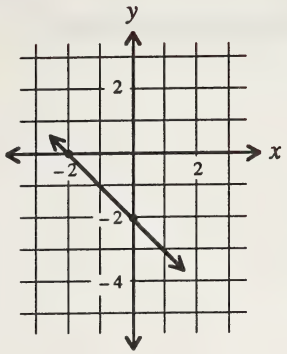
The second point is $(4, 3)$.

$m=\frac{y_2-y_1}{x_2-x_1}$

$m=\frac{3-2}{4-0}$

$m=\frac{1}{4}$

b.



The first point is $(-2, 0)$.

The second point is $(0, -2)$.

$$m = \frac{y_1 - y_2}{x_1 - x_2}$$

$$m = \frac{0-(-2)}{(-2)-0}$$

$$m = -1$$

7. a. $\frac{12}{25} \times 100\% = 0.48 \times 100\%$

$= 48\%$

b. $\frac{17}{20} \times 100\% = 0.85 \times 100\%$

$= 85\%$

c. $\frac{9}{60} \times 100\% = 0.15 \times 100\%$

$= 15\%$

Exploring Topic 1

Activity 1

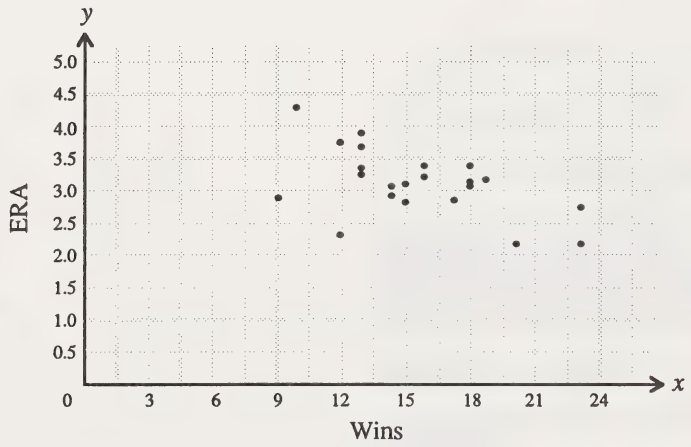
Plot sets of bivariate data on a scatterplot.

1. a. Volume is independent and mass is dependent.

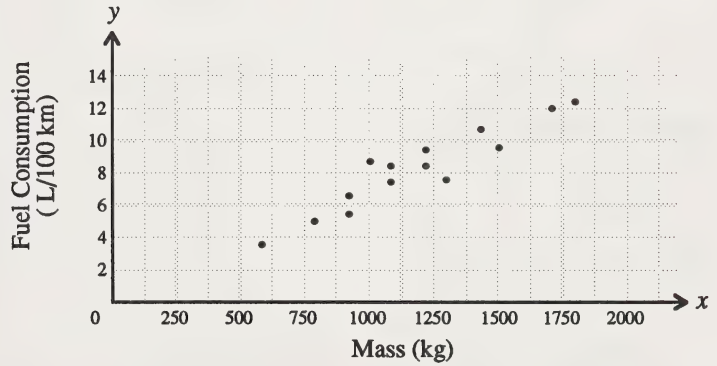
   b. no relation

   c. Time is independent and distance travelled is dependent.

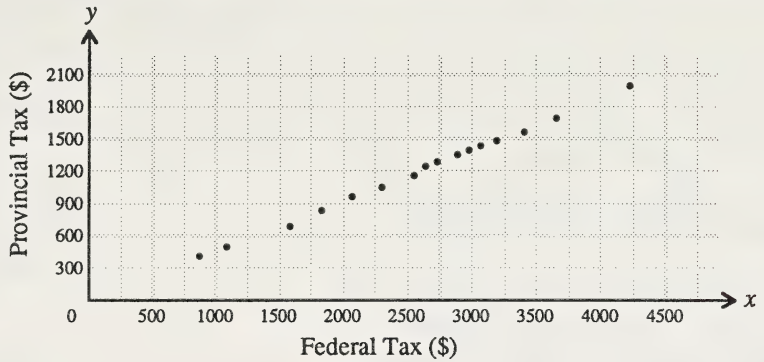
2. Wins versus ERA for Distance League Starting Pitchers, 1989
(ERA means earned run average.)



3. Mass versus Fuel Consumption of Vehicles



4. Federal and Provincial Income Tax Payable for Northland Before Surtax, 1989



Activity 2

Plot a line of best fit on a scatterplot using the median fit method.

1. There are twenty-two points; thus, there will be seven points in each of the first and third strips. The second strip will have eight points.

   The first strip has the following points:
   (9, 2.92), (10, 4.29), (12, 2.44), (12, 3.69), (13, 3.25),
   (13, 3.32), and (13, 3.67)

   The median for the x-coordinate is as follows:
   9, 10, 12, 12, 13, 13, 13
   → 12

The median for the y-coordinate is as follows:
2.44, 2.92, 3.25, 3.32, 3.67, 3.69, 4.29
→ 3.32

The second strip has the following points:
(13, 3.86), (14, 2.92), (14, 3.14), (15, 2.72), (15, 3.08), (16, 3.26), (16, 3.43), and (17, 2.91)

The median for the x-coordinate is as follows:
13, 14, 14, 15, 15, 16, 16, 17
→ $\frac{15+15}{2}=15$

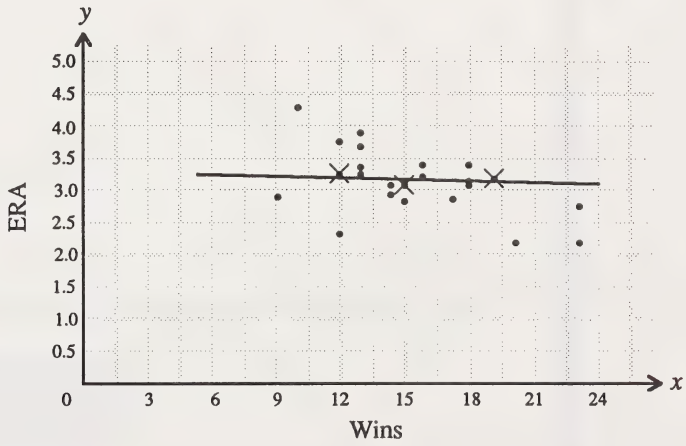
The median for the y-coordinate is as follows:
2.72, 2.91, 2.92, 3.08, 3.14, 3.26, 3.43, 3.86
→ $\frac{3.08+3.14}{2}=3.11$

The third strip has the following points:
(18, 3.18), (18, 3.19), (18, 3.41), (19, 3.12), (20, 2.22), (23, 2.26), and (23, 2.73)

The median for the x-coordinate is as follows:
18, 18, 18, 19, 20, 23, 23
→ 19

The median for the y-coordinate is as follows:
2.22, 2.26, 2.73, 3.12, 3.18, 3.19, 3.41
→ 3.12

Wins versus ERA for Distance League Starting Pitchers, 1989
(ERA means earned run average.)



2. The first strip has the following points:
(600, 3.9), (800, 5.0), (900, 5.6), (900, 6.2), and (1000, 8.5)

The median for the x-coordinate is as follows:
600, 800, 900, 900, 1000
→ 900

The median for the y-coordinate is as follows:
3.9, 5.0, 5.6, 6.2, 8.5
→ 5.6

The second strip has the following points:
(1100, 7.5), (1100, 8.2), (1200, 8.2), and (1200, 9.6)

The median for the x-coordinate is as follows:
1100, 1100, 1200, 1200

$$\rightarrow \frac{1100+1200}{2} = 1150$$

The median for the y-coordinate is as follows:
7.5, 8.2, 8.2, 9.6

$$\rightarrow \frac{8.2+8.2}{2} = 8.2$$

The third strip has the following points:
(1300, 7.8), (1400, 10.4), (1500, 9.8), (1700, 12.0), and (1800, 12.3)

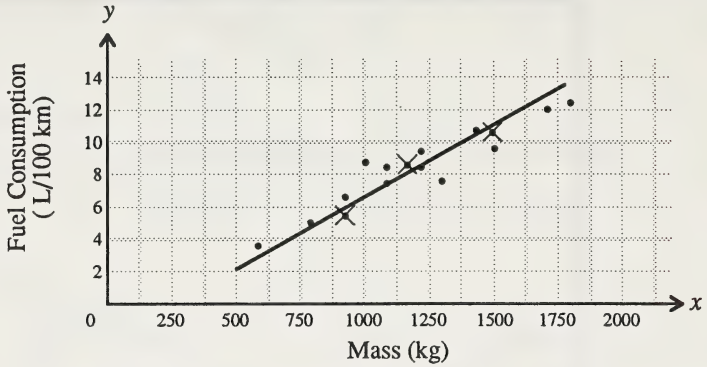
The median for the x-coordinate is as follows:
1300, 1400, 1500, 1700, 1800

$$\rightarrow 1500$$

The median for the y-coordinate is as follows:
7.8, 9.8, 10.4, 12.0, 12.3

$$\rightarrow 10.4$$

Mass versus Fuel Consumption of Vehicles



3. The first strip has the following points:
(820, 381.20), (1060, 493.60), (1540, 716.40), (1780, 828.80), and (2040, 949.90)

The median for the x-coordinate is as follows:
820, 1060, 1540, 1780, 2040

$$\rightarrow 1540$$

The median for the y-coordinate is as follows:
381.20, 493.60, 716.40, 828.80, 949.90

$$\rightarrow 716.40$$

The second strip has the following points:
(2280, 1060.30), (2560, 1190.10), (2620, 1218.20), (2740, 1274.40), (2860, 1330.60), and (2980, 1386.80)

The median for the x-coordinate is as follows:
2280, 2560, 2620, 2740, 2860, 2980

$$\rightarrow \frac{2620+2740}{2} = 2680$$

The median for the y-coordinate is as follows:
1060.30, 1190.10, 1218.20, 1274.40, 1330.60, 1386.80

$$\frac{1218.20+1274.40}{2}=1246.30$$

The third strip has the following points:
(3020, 1404.20), (3160, 1469.10), (3340, 1553.40),
(3660, 1702.60), and (4240, 1972.90)

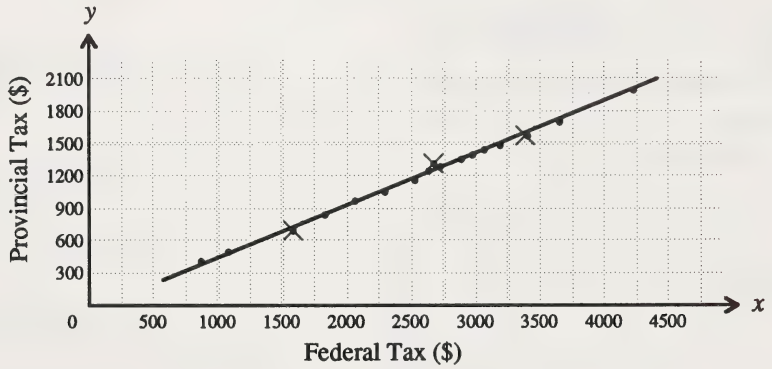
The median for the x-coordinate is as follows:
3020, 3160, 3340, 3660, 4240

→ 3340

The median for the y-coordinate is as follows:
1404.20, 1469.10, 1553.40, 1702.60, 1972.90

→ 1553.40

Federal and Provincial Income Tax Payable in Northland Before Surtax, 1989



Activity 3

Use the equation of the line of best fit to generate new data for a population.

1. a. i. The fuel consumption is about 4.5 L/100 km.

ii. The fuel consumtion is about 9.0 L/100 km.

iii. The fuel consumption is about 13.0 L/100 km.

To answer the next two questions, first you must find the equation of the line of best fit.

$$m = \frac{y_1 - y_2}{x_1 - x_2}$$

$$m = \frac{13.0 - 4.5}{1750 - 750}$$

$$m = \frac{8.5}{1000} \text{ or } 0.0085$$

$$y = mx + b$$

$$y = 0.0085x + b$$

$$9.0 = 0.0085(1275) + b$$

$$9.0 = 10.8375 + b$$

$$-1.8375 = b$$

$$y = 0.0085x - 1.8375$$

iv. $y = 0.0085x - 1.8375$

$y = 0.0085(400) - 1.8375$

$y = 3.4 - 1.8375$

$y = 1.5625$

The fuel consumption is about 1.56 L/100 km.

v. $y = 0.0085x - 1.8375$

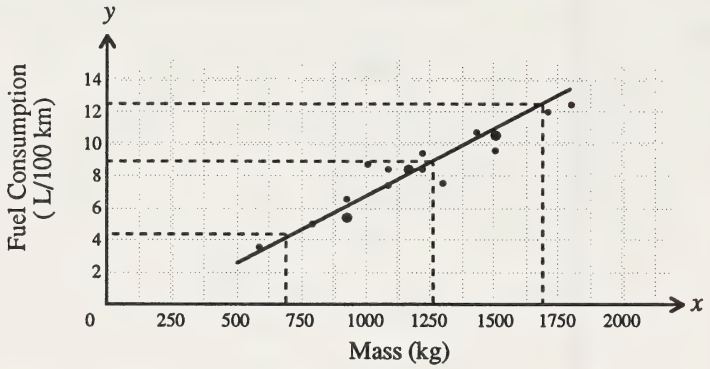
$y = 0.0085(4500) - 1.8375$

$y = 38.25 - 1.8375$

$y = 36.4125$

The fuel consumption is about 36.41 L/100 km.

Mass versus Fuel Consumption of Vehicles



b. i. The mass is about 700 kg.

ii. The mass is about 1270 kg.

iii. The mass is about 1700 kg.

iv. $y = 0.0085x - 1.8375$

$1.2 = 0.0085x - 1.8375$

$3.0375 = 0.0085x$

$357.4 \doteq x$

The mass is about 357 kg.

v. $y = 0.0085x - 1.8375$

$17.9 = 0.0085x - 1.8375$

$19.7375 = 0.0085x$

$2322.1 \doteq x$

The mass is about 2322 kg.

2. a. i. The pitcher will have 8 wins.

ii. The pitcher will have 13 wins.

iii. The pitcher will have 18 wins.

To answer the next two questions, you must first find the equation of the line of best fit.

$m = \dfrac{y_1 - y_2}{x_1 - x_2}$

$m = \dfrac{3.40 - 3.00}{8 - 18}$

$m = \dfrac{0.40}{-10}$ or -0.04

$y = mx + b$

$y = -0.04x + b$

$3.2 = -0.04(13) + b$

$3.2 = -0.52 + b$

$3.72 = b$

$y = -0.04x + 3.72$

iv. $y = -0.04x + 3.72$

$3.60 = -0.04x + 3.72$

$-0.12 = -0.04x$

$3 = x$

The pitcher will have 3 wins.

v. $y = -0.04x + 3.72$

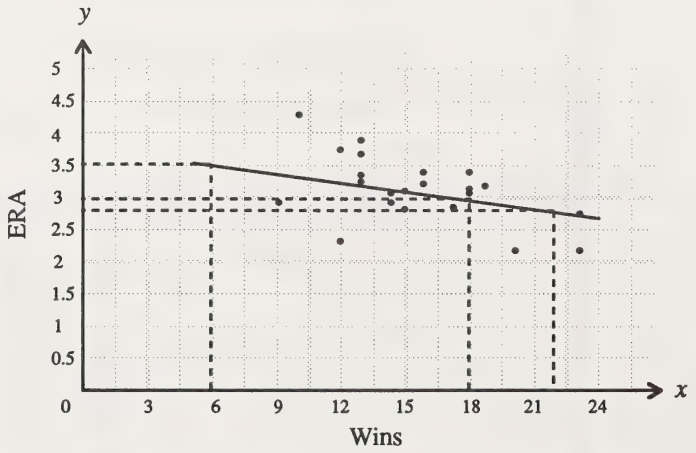
$1.90 = -0.04x + 3.72$

$-1.82 = -0.04x$

$45.5 = x$

The pitcher will have about 46 wins.

Wins versus ERA for Distance League Starting Pitchers, 1989
(ERA means earned run average.)



b. i. The pitcher's ERA is about 3.50.

ii. The pitcher's ERA is about 2.90.

iii. The pitcher's ERA is about 2.75.

iv. $y = -0.04x + 3.72$

$y = -0.04(1) + 3.72$

$y = -0.04 + 3.72$

$y = 3.68$

The pitcher's ERA is 3.68.

v. $y = -0.04x + 3.72$

$y = -0.04(30) + 3.72$

$y = -1.2 + 3.72$

$y = 2.52$

The pitcher's ERA is 2.52.

3. a. i. The federal tax will be about \$1350

ii. The federal tax will be about \$2000

iii. The federal tax will be about \$3950

To answer the next two questions, first you must find the equation of the line of best fit.

$$(y - y_1) = \frac{y_1 - y_2}{x_1 - x_2}(x - x_1)$$

$$y - 900 = \frac{1800 - 600}{3950 - 1350}(x - 2000)$$

$$y - 900 = \frac{1200}{2600}(x - 2000)$$

$$y - 900 = \frac{6}{13}(x - 2000)$$

$$13y - 11\,700 = 6x - 12\,000$$

$$0 = 6x - 13y - 300$$

iv. $0 = 6x - 13y - 300$

$0 = 6x - 13(2200) - 300$

$0 = 6x - 28\,600 - 300$

$0 = 6x - 28\,900$

$28\,900 = 6x$

$4816.67 \doteq x$

The federal tax will be about \$4816.67.

v. $0 = 6x - 13y - 300$

$0 = 6x - 13(3500) - 300$

$0 = 6x - 45\,500 - 300$

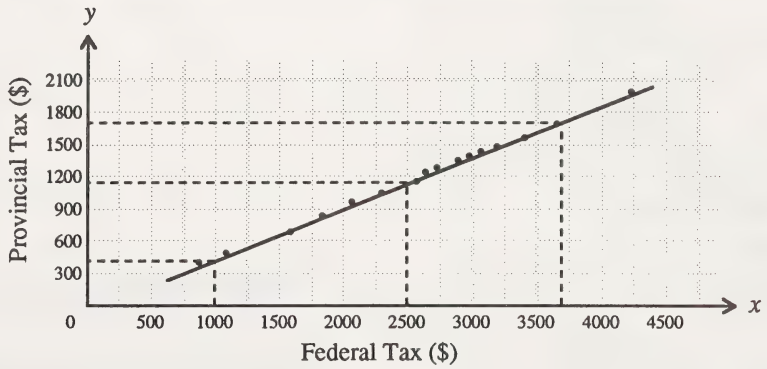
$0 = 6x - 45\,800$

$45\,800 = 6x$

$7633.33 \doteq x$

The federal tax will be about \$7633.33.

Federal and Provincial Income Tax Payable in Northland Before Surtax, 1989



b. i. The provincial tax will be about \$450.

ii. The provincial tax will be about \$1150.

iii. The provincial tax will be about \$1700.

iv.
$$0 = 6x - 13y - 300$$
$$0 = 6(300) - 13y - 300$$
$$0 = 1800 - 13y - 300$$
$$0 = 1500 - 13y$$
$$-1500 = -13y$$
$$115.38 \doteq y$$

The provincial tax will be about \$115.38.

v.
$$0 = 6x - 13y - 300$$
$$0 = 6(5500) - 13y - 300$$
$$0 = 33\,000 - 13y - 300$$
$$0 = 32\,700 - 13y$$
$$-32\,700 = -13y$$
$$2515.38 \doteq y$$

The provincial tax will be about \$2515.38.

Activity 4

Determine the strength and type of correlation between the variables of a bivariate distribution.

1. $r \doteq 1 - \frac{1}{3}$

   $\doteq \frac{2}{3}$

   $\doteq 0.66$

2. $r \doteq -\left(1 - \frac{2}{3}\right)$

   $\doteq -\frac{1}{3}$

   $\doteq -0.333$

3. $r = 0$

Activity 5

Collect, organize, and analyze sets of bivariate data.

The following solution is only an example of the components that you should have in your solution. Your solution should not be the same as this example.

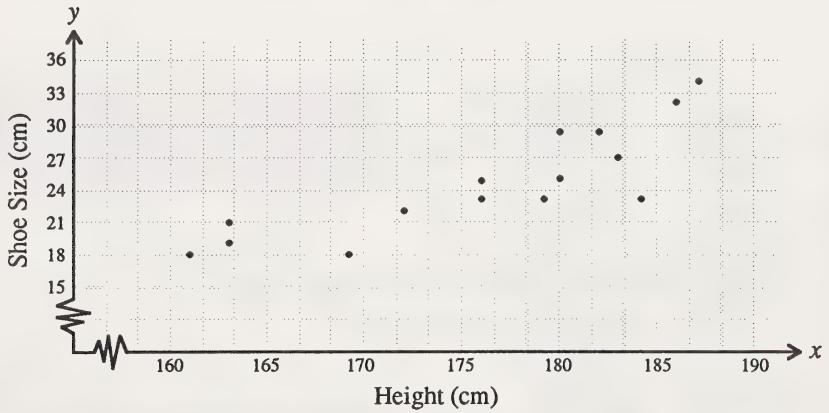
Solutions to the other three questions will be similar to the solution shown for question 1. Since answers will vary, solutions for questions 2, 3, and 4 are not provided.

1. The chart of collected data is as follows:

Person	Height	Shoe Size
Eugene A.	180 cm	25 cm
Angie B.	163 cm	19 cm
Bobby C.	182 cm	29 cm
Debbie D.	176 cm	23 cm
Charlie E.	186 cm	32 cm
Tanya F.	163 cm	21 cm
Drago G.	176 cm	25 cm
Michele H.	161 cm	18 cm
Doug I.	180 cm	29 cm
Mai J.	172 cm	22 cm
Joe K.	184 cm	23 cm
Natalia L.	179 cm	23 cm
Kwok M.	183 cm	27 cm
Jolene N.	169 cm	18 cm
Guy O.	187 cm	34 cm

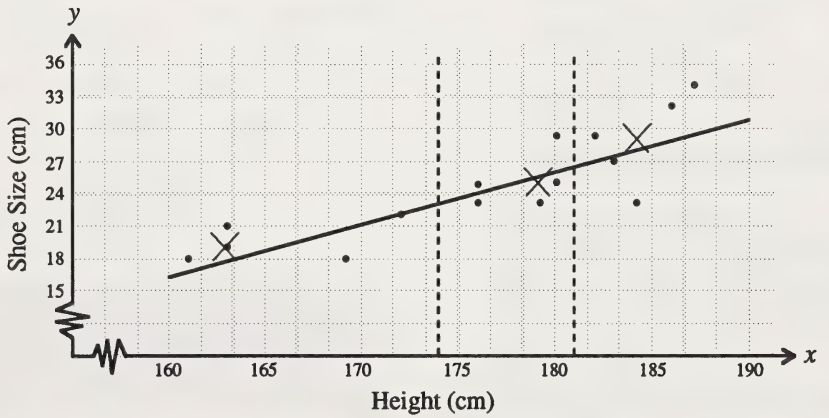
The scatterplot of the data is as follows:

Height versus Shoe Size

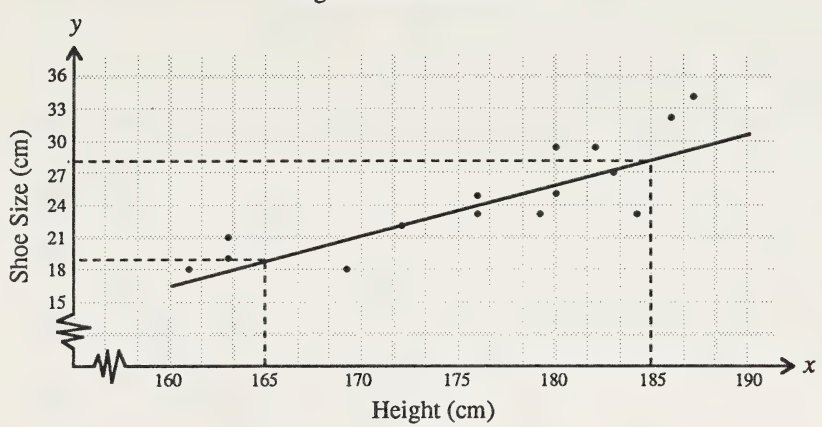


The line of best fit is as follows:

Height versus Shoe Size



The equation of the line of best fit is as follows:



From the graph, when the height is 185 cm, the shoe size is 28 cm. When the height is 165 cm, the shoe size is 19 cm. Use the points (185, 28) and (165, 19) to obtain the equation of the line.

$$y - y_2 = \frac{y_1 - y_2}{x_1 - x_2}(x - x_2)$$

$$y - 19 = \frac{28 - 19}{185 - 165}(x - 165)$$

$$y - 19 = \frac{9}{20}(x - 165)$$

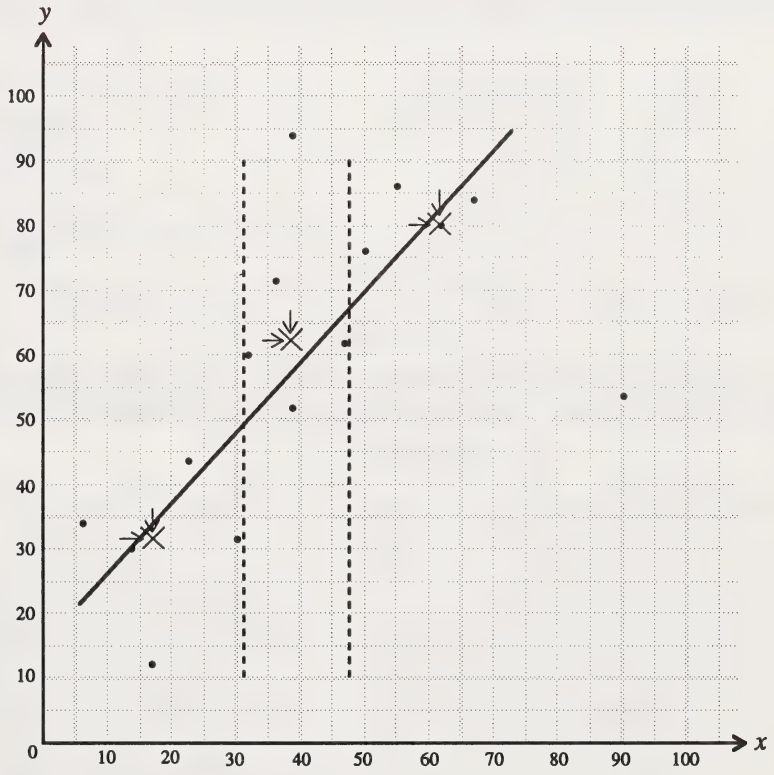
$$20y - 380 = 9x - 1485$$

$$0 = 9x - 20y - 1105$$

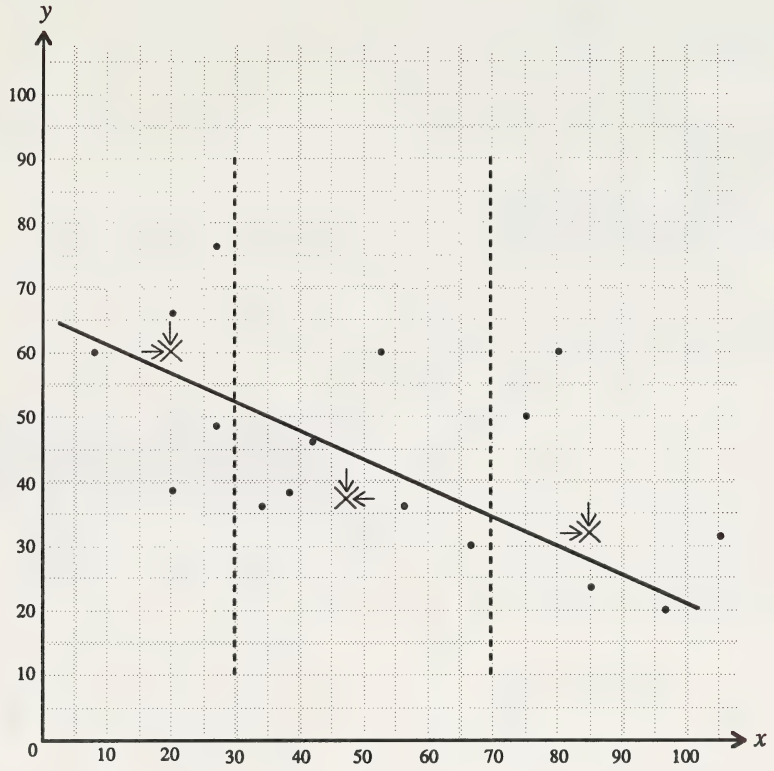
There is a definite positive correlation between the height of a person and the shoe size of that person. Neither of the factors is considered to be independent. Both factors are considered to be dependent on a third factor.

Extra Help

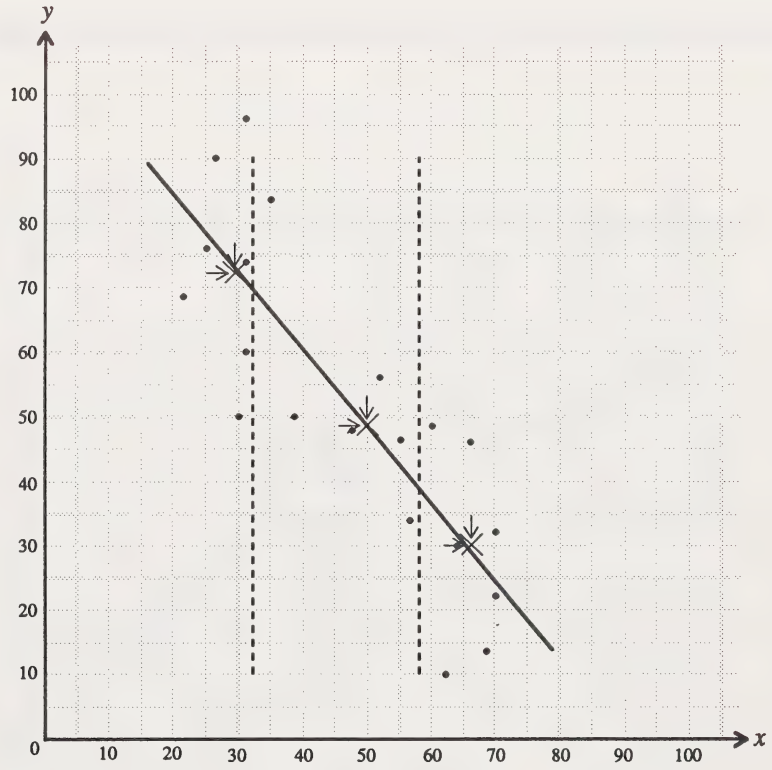
1.



2.



3.



Extensions

1. -0.531
2. 0.952
3. 0.999 999

Exploring Topic 2

Activity 1

Design and administer a yes/no simple survey and collect and organize the results of the survey.

1. Answers will vary. The following are only some of many possible solutions.
   a. Do you prefer to watch hockey over baseball?
   b. Do you prefer a large breakfast in the morning?
   c. Do you prefer to drive large cars?
   d. Would you rather vacation in Europe than on a Pacific island?
2. Answers will vary. The following are only some of many possible solutions.
   a. Do you prefer to eat bananas over oranges?
      Do you prefer to eat oranges over apples?
      Do you prefer to eat apples over bananas?
   b. Would you prefer Brian Mulroney over Audrey McLaughlin to be the next prime minister?
      Would you prefer Audrey McLaughlin over Jean Chrétien to be the next prime minister?
      Would you prefer Brian Mulroney over Jean Chrétien to be the next prime minister?
3. Your answers will be near the following. The following are only some of many possible solutions.
   a. 4 heads
   b. The probability of getting 2 heads is $\frac{7}{64}$ or 0.109 375.
      The probability of getting 3 heads is $\frac{7}{32}$ or 0.218 75.
      The probability of getting 7 heads is $\frac{1}{32}$ or 0.031 25.
      The probability of getting 8 heads is $\frac{1}{256}$ or 0.003 906 25.
   c. $\frac{93}{256}$ or 0.363 281

4. Answers will vary. The following is only one of many possible solutions.

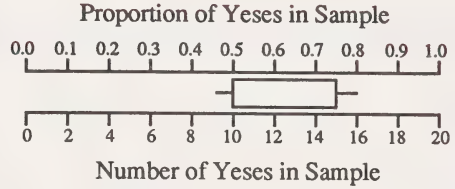
   The population percentage is $14 \div 20 = 0.7$ or 70% and the sample size is 20. Use a random number table selecting two digits at a time. If the selected number is 01 to 70, the response is yes. If the selected number is 00 or 71 to 99, the response is no. Fifty sets of twenty responses will be gathered.

Activity 2

Draw box and whisker plots of the results of multiple samples.

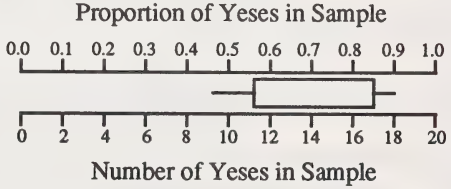
1. There can be a maximum of three values in the left whisker since 5% of 60 equals 3. The nine yeses have a frequency of 1; thus, start the left end of the left whisker at 9. The left side of the box has to start at ten yeses because the frequency for the ten yeses is 5, and these five values cannot be split. These five values cannot be placed in the left whisker, since the whisker can hold a maximum of three values and it already has one value. The frequency for fifteen yeses is 6 and the frequency for sixteen yeses is 2; thus, draw the right side of the box at fifteen yeses. Draw the right end of the right whisker at 16.

   This box and whisker plot has one value in the left whisker, fifty-seven values in the box, and two values in the right whisker.



2. The left whisker can have a maximum of five values since 5% of 100 equals 5. The nine yeses have a frequency of 1 and the ten yeses have a frequency of 4. Thus, these five values can be shown in the left whisker. Start the end of the left whisker at nine yeses and start the left side of the box at eleven yeses. The seventeen yeses have a frequency of 4 and the eighteen yeses have a frequency of 3. The three values for the eighteen yeses can be placed in the right whisker since it can hold a maximum of five values. The right side of the box ends at 17 and the right end of the right whisker ends at 18.

   This box and whisker plot has five values in the left whisker, ninety-two values in the box, and three values in the right whisker.



Activity 3

Use a 90% box and whisker plot chart to find the confidence interval for a survey result.

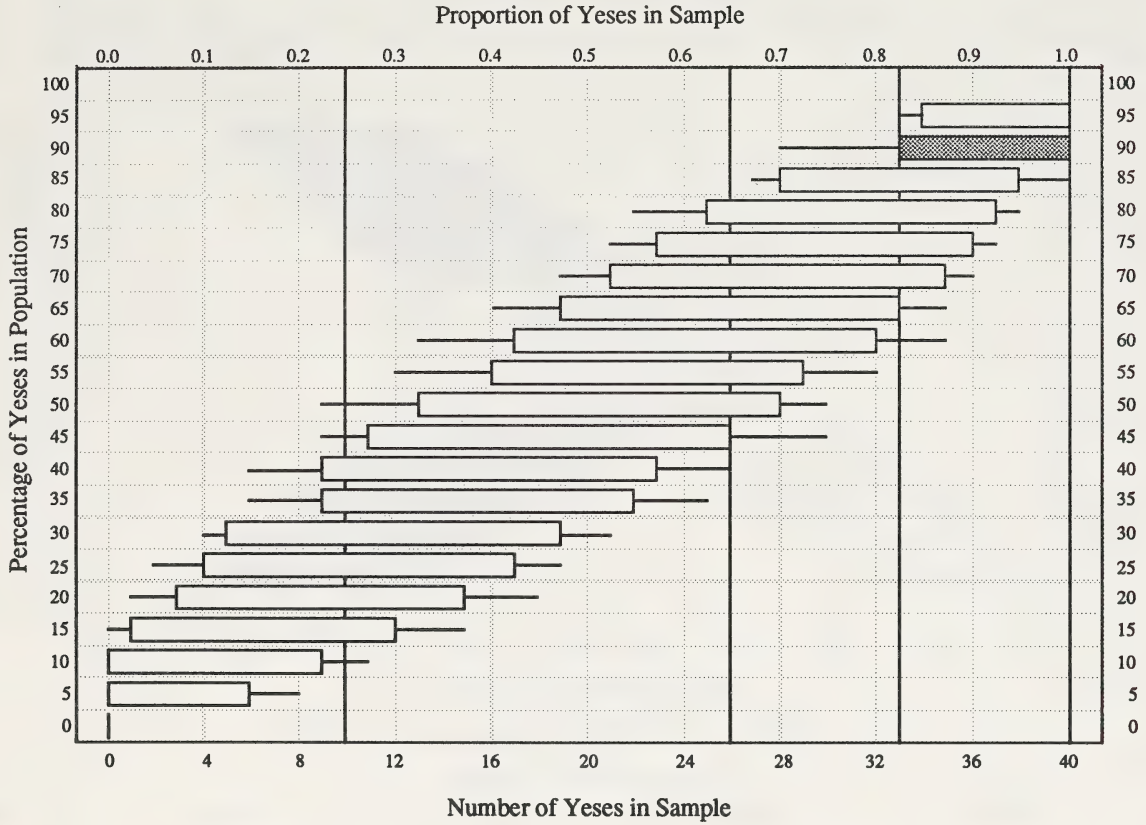
1. a. A sample proportion of 0.25 is unlikely.

b. A sample proportion of 0.65 is unlikely.

c. A sample proportion of 0.825 is likely.

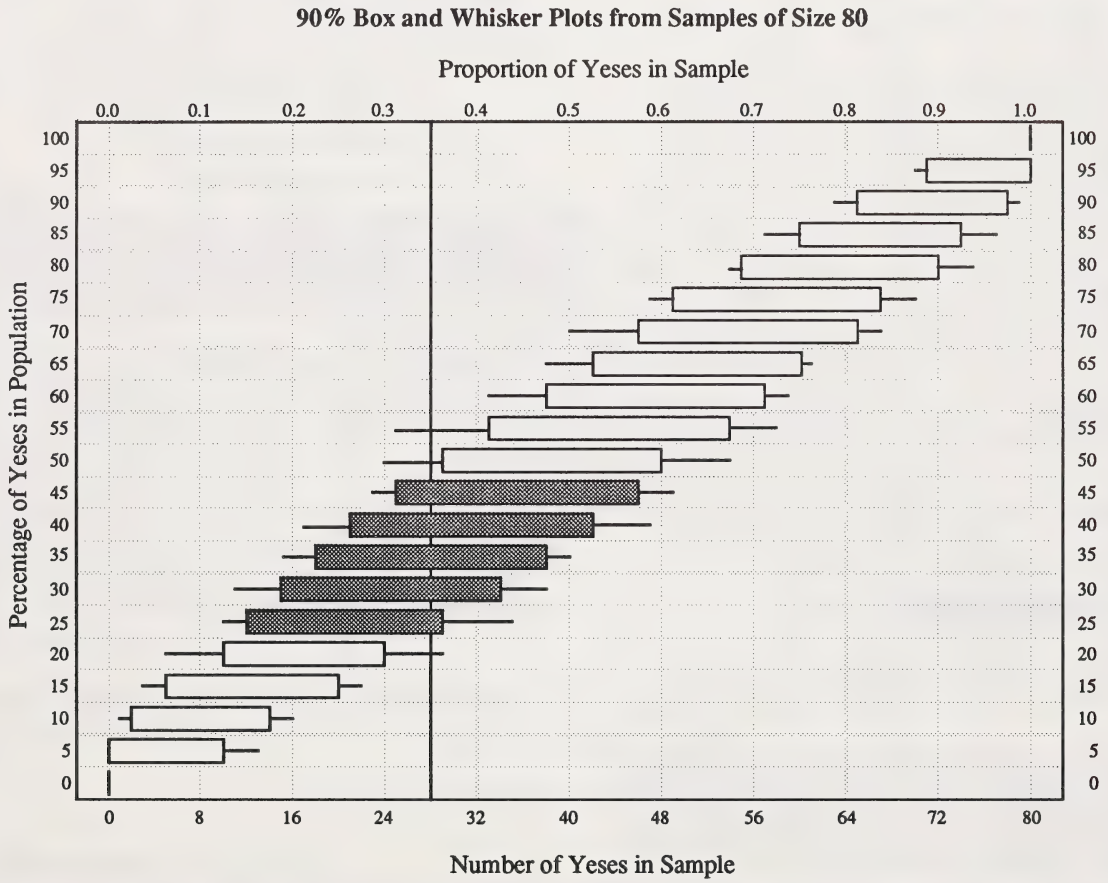
d. A sample proportion of 1.00 is likely.

90% Box and Whisker Plots from Samples of Size 40

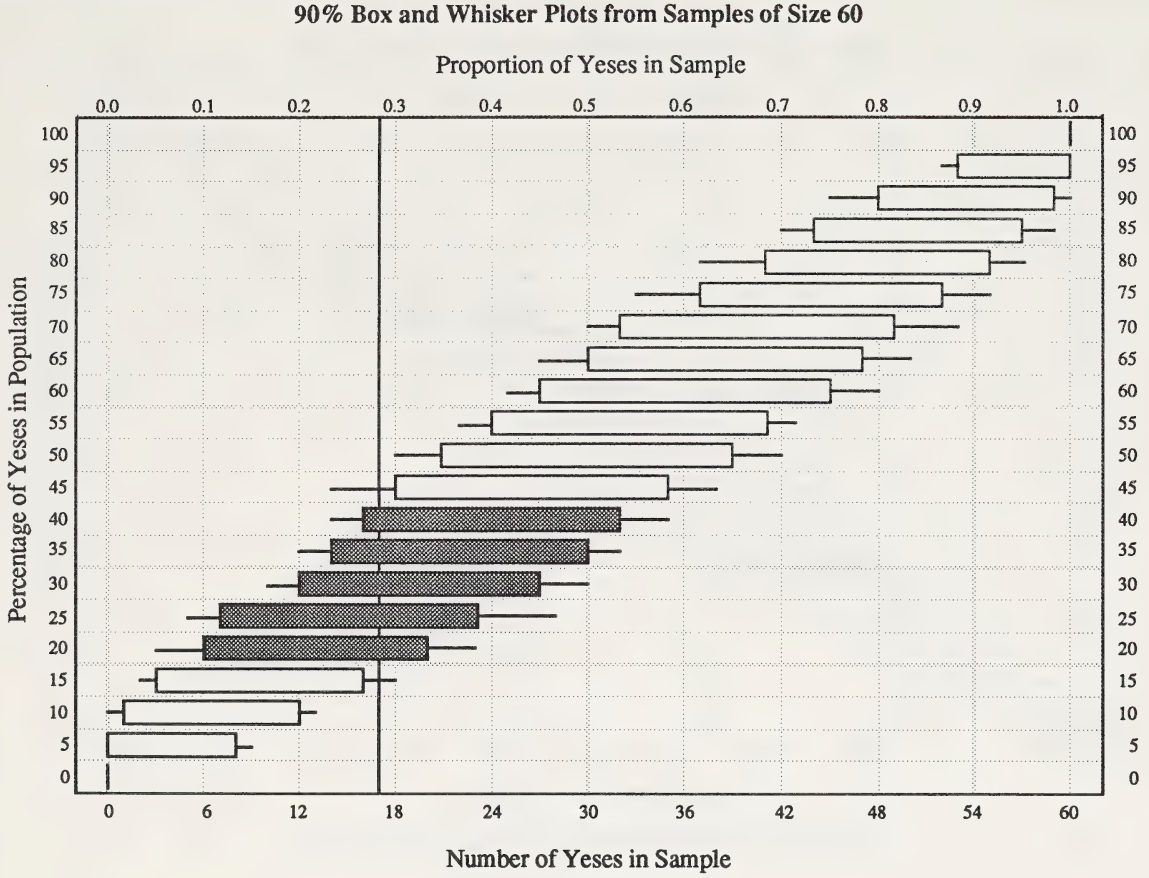


2. a. unlikely b. unlikely c. likely d. likely e. likely

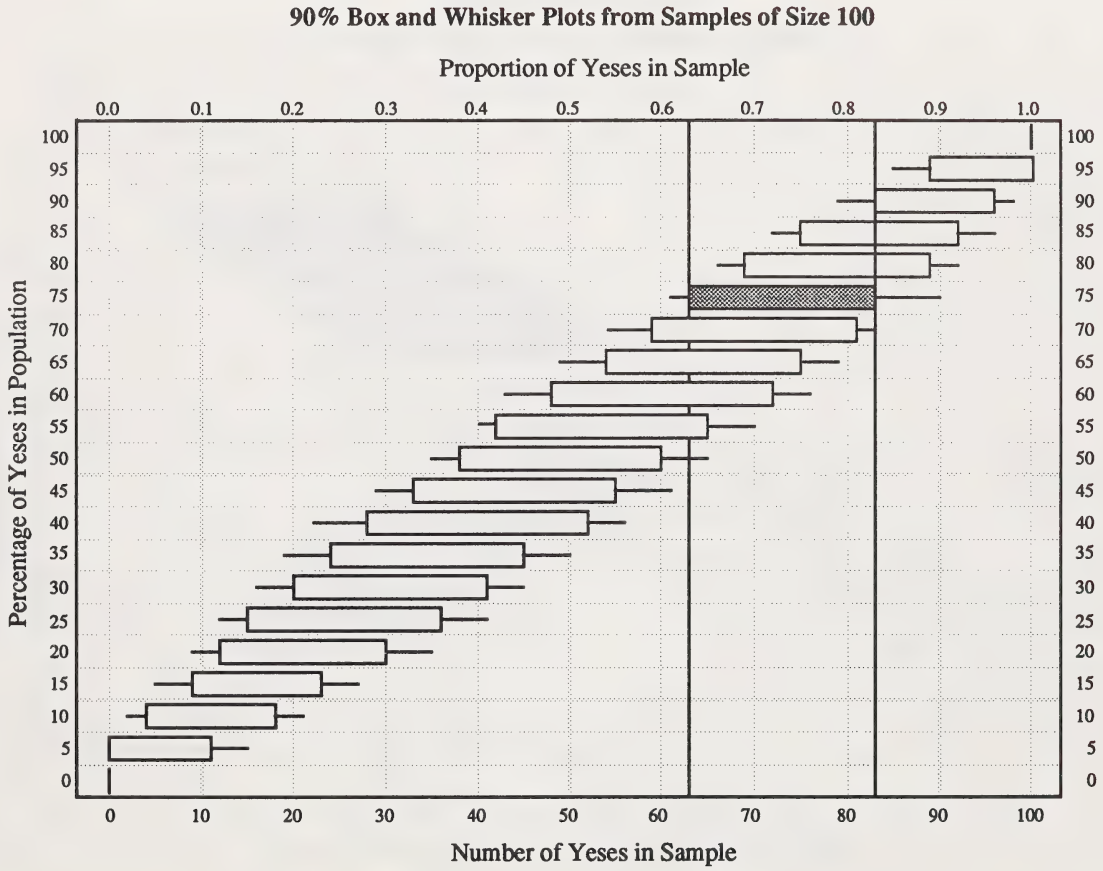
f. likely g. likely h. unlikely i. unlikely j. unlikely



3. This is likely for the population percentages of 20% to 40% inclusive.

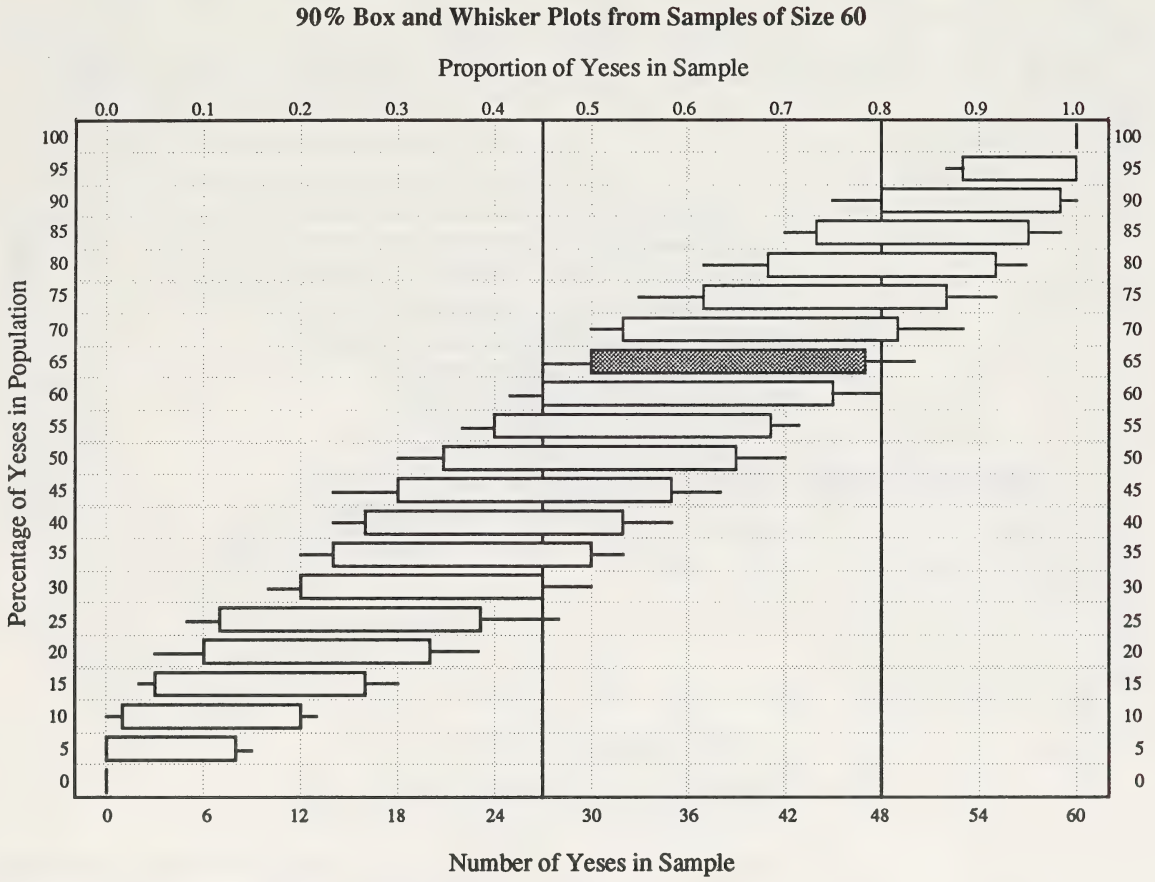


4. The sample proportions will go from 0.63 to 0.83 inclusive.



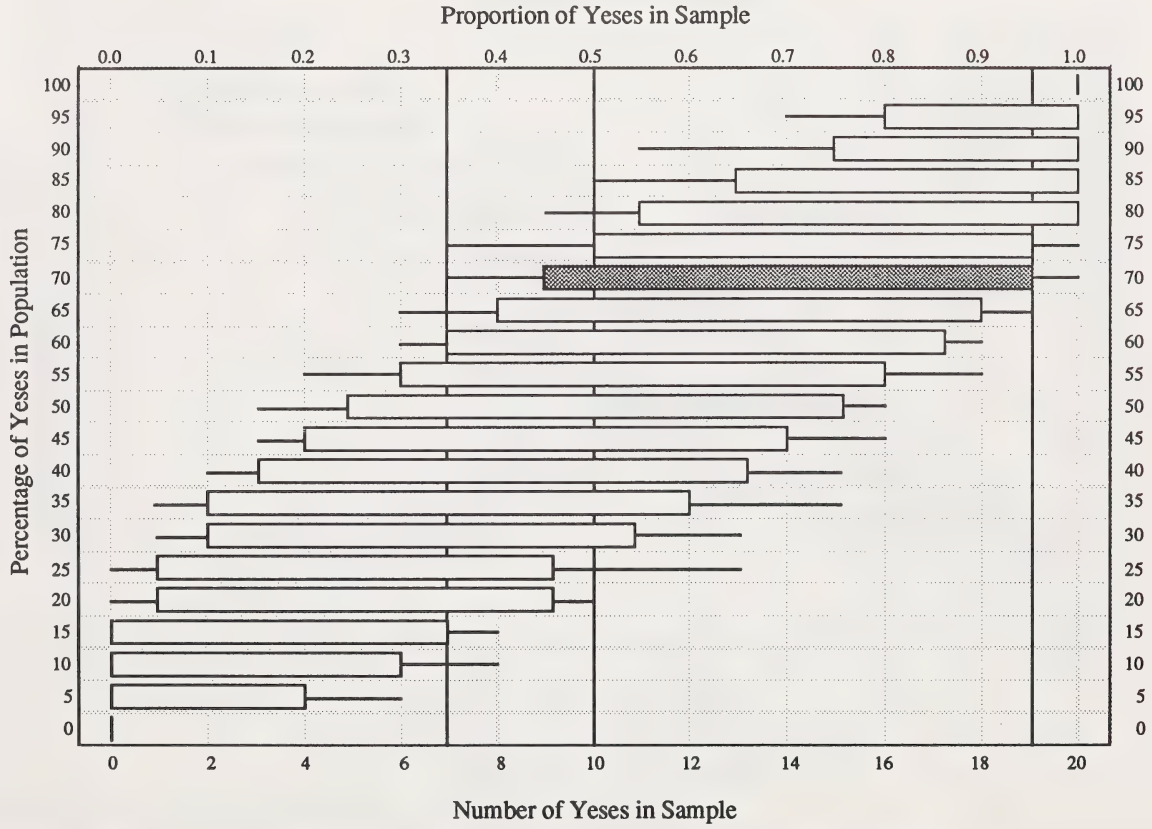
5. a. A sample proportion of 0.45 is unlikely.

b. A sample with forty-eight people having alcohol in their blood is unlikely.

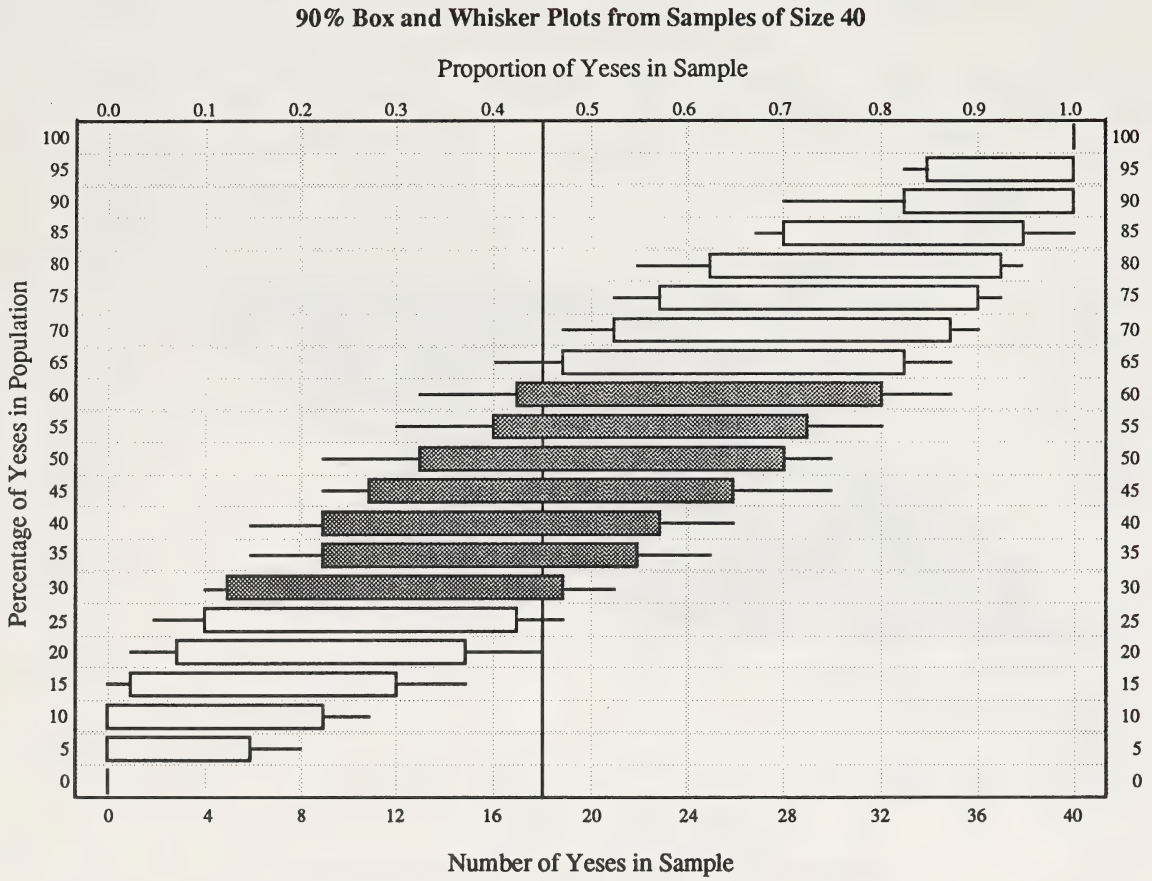


6. a. It is likely that a sample proportion of 0.95 of the people will approve of the tax.
   b. It is likely that ten people will approve of the tax.
   c. It is unlikely that a sample proportion of 0.35 of the people will approve of the tax.

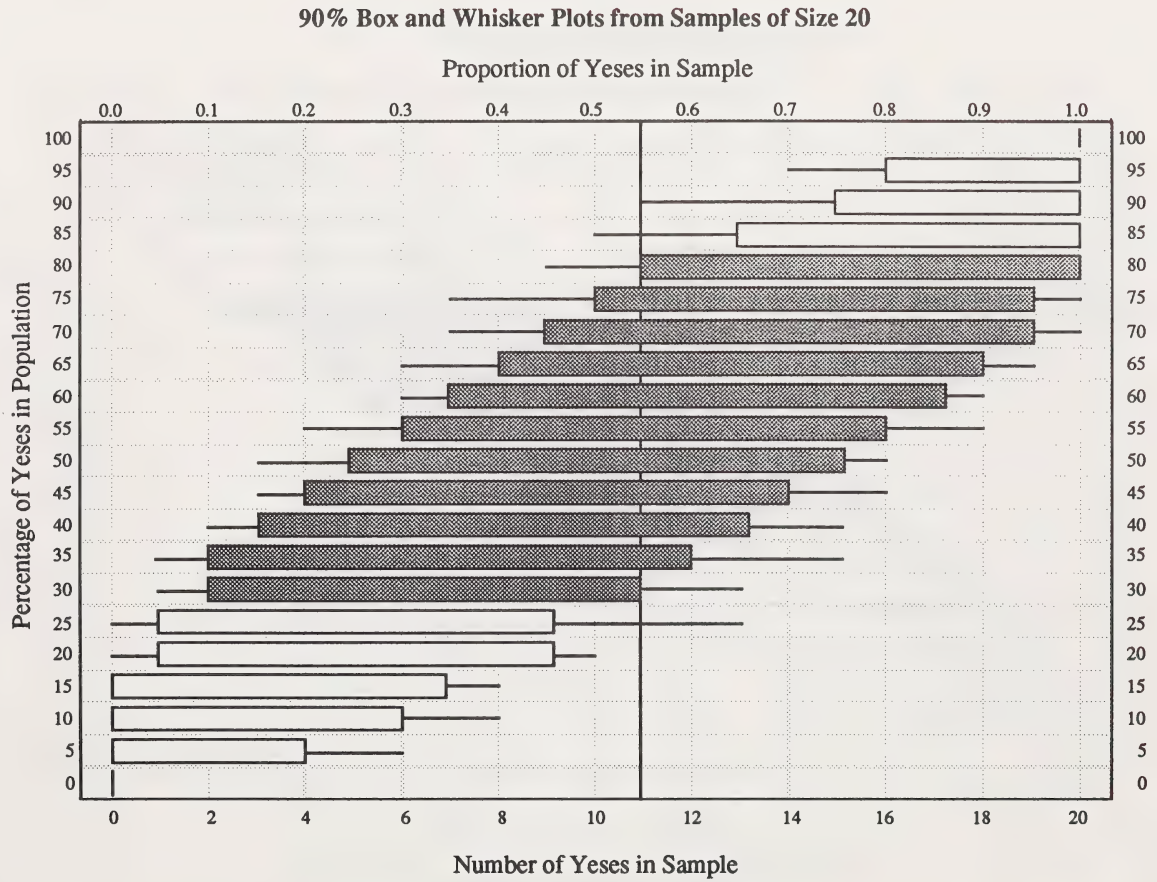
90% Box and Whisker Plots from Samples of Size 20



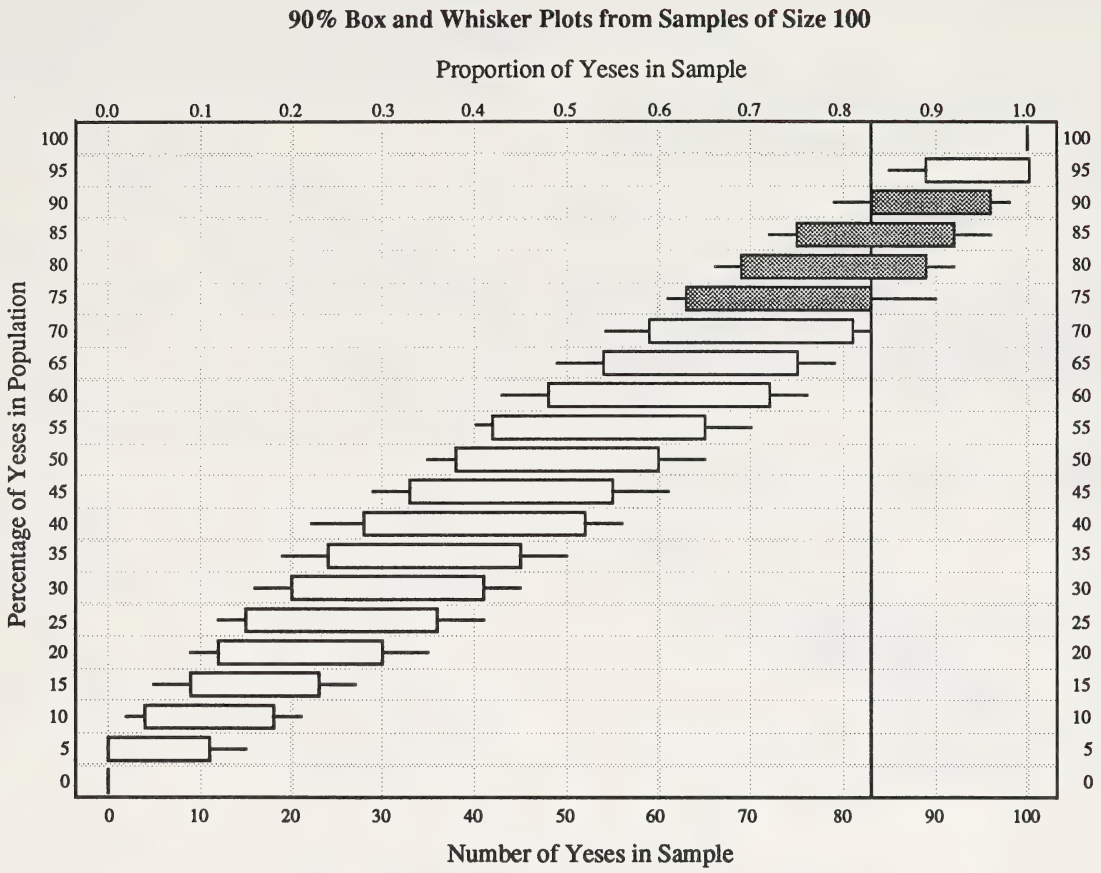
7. Thirty percent to sixty percent of the people will buy the car.



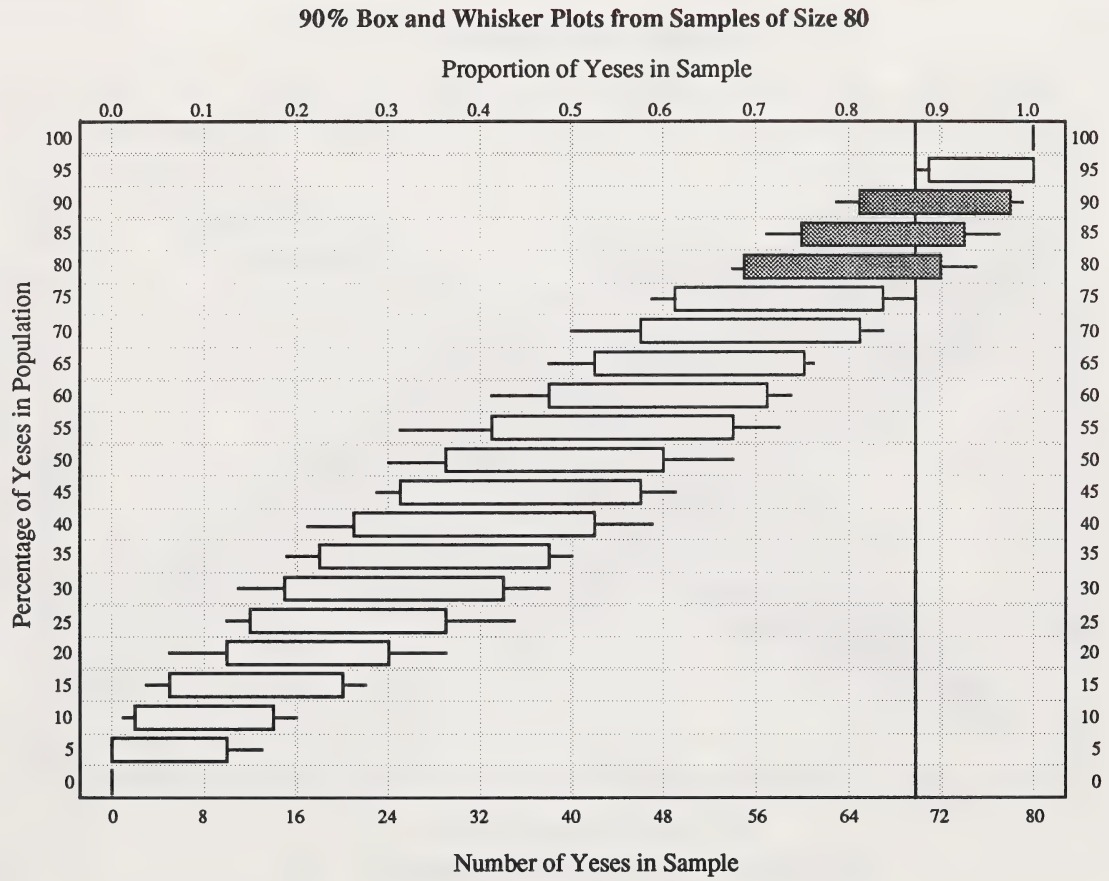
8. a. The confidence interval is from 30% to 80% inclusive.



b. The confidence interval is from 75% to 90% inclusive.



c. The confidence interval is from 80% to 90% inclusive.



Activity 4

Draw statistical conclusions and make inferences to populations, and explain the confidence with which such conclusions and inferences are made based on the results of yes/no surveys.

1. 0.70 is not a likely sample proportion.

2. Ninety percent of the samples will be likely.
   $90\% \text{ of } 500 = 0.90 \times 500$
   $= 450$
   Four hundred fifty of the population percentages will be likely.

3. The sample size should be decreased.

4. From the box and whisker plots for a sample size of 60, the 90% confidence interval will be 70% to 90%. A vertical line from 49 intersects the 70%, 75%, 80%, 85%, and 90% boxes.

   Seventy percent of 60 is 42 and 90% of 60 is 54. If the sixty people of a random sample of this population were asked if they ate breakfast every morning, then 42 to 54 people would respond yes. These results will be true for 90 samples out of every 100 samples.

5. From the box and whisker plots for a sample size of 40, the 90% confidence interval will be 25% to 60%. A vertical line from 17 intersects the 25%, 30%, 35%, 40%, 45%, 50%, 55%, and 60% boxes. Twenty-five percent of 40 is 10 and 60% of 40 is 24.

   If the forty people of a random sample of this population were asked if they favoured running the public school year-round, then 10 to 24 people would respond yes. These results will be true for 90 samples out of every 100 samples.

6. From the box and whisker plots for a sample size of 80, the 90% confidence interval will be 85% to 95%. A vertical line from 73 intersects the 85%, 90%, and 95% boxes. Eighty-five percent of 80 is 68 and 95% of 80 is 76.

   If the eighty people of a random sample of this population were asked if they wanted more parks, 68 to 76 people would respond yes. The results will be true for 90 samples out of every 100 samples.

Activity 5

Assess the strengths, weaknesses, and biases of given samples.

1. a. Yes. Assuming that every student has a phone and that the phone numbers are assigned randomly, this is a random sampling of students.

   b. Yes.

   c. No. Students will not be selected independently. Entire families may be selected.

2. This situation may or may not be a random selection. It is impossible to tell from the results of the selection whether a random selection has been performed.

3. It would be impossible to get a complete and up-to-date list of all of the people that live in Calgary. Since you will not be able to list all of the people that live in Calgary, all of the people will not have the same chance of being selected.

4. a. self-selected

   b. stratified random

   c. judgement

5. Answers will vary. Some possible solutions are as follows:

   - People may be giving answers that they believe are socially acceptable.
   - People may be trying to give the answers that they believe the interviewer wants to hear.
   - People may be trying to seem knowledgeable by trying to answer questions that they do not understand.
   - People often do not correctly remember numbers.

6. Answers will vary. Some possible solutions are as follows:

   - People in the armed forces or in institutions may not have been surveyed.
   - Women may consider it more acceptable to be considered married and thus, report themselves as being married.

7. Answers will vary.

Extra Help

1. Note that the diagram is at the end of question 5.

Number of Yeses	Sample Proportion	Frequency	Proportion of All Trials	
0	0.00	55	0.055	
1	0.05	149	0.149	
2	0.10	166	0.166	
3	0.15	215	0.215	Box
4	0.20	169	0.169	
5	0.25	132	0.132	
6	0.30	59	0.059	
7	0.35	47	0.047	
8	0.40	8	0.008	Right Whisker
9	0.45	0	0.000	
.	.	.	.	
.	.	.	.	
.	.	.	.	
20	1.00	0	0.000	
Total		1000	1.000	

2.

Number of Yeses	Sample Proportion	Frequency	Proportion of All Trials	
0	0.00	17	0.017	Left Whisker
1	0.05	59	0.059	
2	0.10	136	0.136	
3	0.15	183	0.183	
4	0.20	170	0.170	
5	0.25	141	0.141	Box
6	0.30	103	0.103	
7	0.35	77	0.077	
8	0.40	63	0.063	
9	0.45	49	0.049	
10	0.50	2	0.002	Right Whisker
11	0.55	0	0.000	
.	.	.	.	
.	.	.	.	
.	.	.	.	
20	1.00	0	0.000	
Total		1000	1.000	

3.

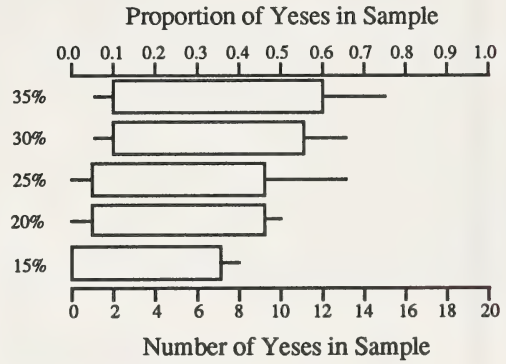
Number of Yeses	Sample Proportion	Frequency	Proportion of All Trials	
0	0.00	4	0.004	} Left Whisker
1	0.05	47	0.047	
2	0.10	82	0.082	
3	0.15	117	0.117	
4	0.20	151	0.151	
5	0.25	157	0.157	} Box
6	0.30	138	0.138	
7	0.35	117	0.117	
8	0.40	101	0.101	
9	0.45	49	0.049	
10	0.50	20	0.020	
11	0.55	11	0.011	} Right Whisker
12	0.60	5	0.005	
13	0.65	1	0.001	
14	0.70	0	0.000	
.	.	.	.	
.	.	.	.	
.	.	.	.	
20	1.00	0	0.000	
Total		1000	1.000	

4.

Number of Yeses	Sample Proportion	Frequency	Proportion of All Trials	
0	0.00	0	0.000	
1	0.05	22	0.022	} Left Whisker
2	0.10	39	0.039	
3	0.15	54	0.054	
4	0.20	93	0.093	
5	0.25	107	0.107	
6	0.30	123	0.123	
7	0.35	140	0.140	} Box
8	0.40	134	0.134	
9	0.45	108	0.108	
10	0.50	83	0.083	
11	0.55	57	0.057	
12	0.60	32	0.032	} Right Whisker
13	0.65	8	0.008	
14	0.70	0	0.000	
.	.	.	.	
.	.	.	.	
.	.	.	.	
20	1.00	0	0.000	
Total		1000	1.000	

5.

Number of Yeses	Sample Proportion	Frequency	Proportion of All Trials	
0	0.00	0	0.000	
1	0.05	7	0.007	} Left Whisker
2	0.10	52	0.052	
3	0.15	78	0.078	
4	0.20	102	0.102	
5	0.25	111	0.111	
6	0.30	123	0.123	
7	0.35	129	0.129	} Box
8	0.40	113	0.113	
9	0.45	98	0.098	
10	0.50	76	0.076	
11	0.55	57	0.057	
12	0.60	42	0.042	
13	0.65	11	0.011	
14	0.70	0	0.000	} Right Whisker
15	0.75	1	0.001	
16	0.80	0	0.000	
.	.	.	.	
.	.	.	.	
.	.	.	.	
20	1.00	0	0.000	
Total		1000	1.000	

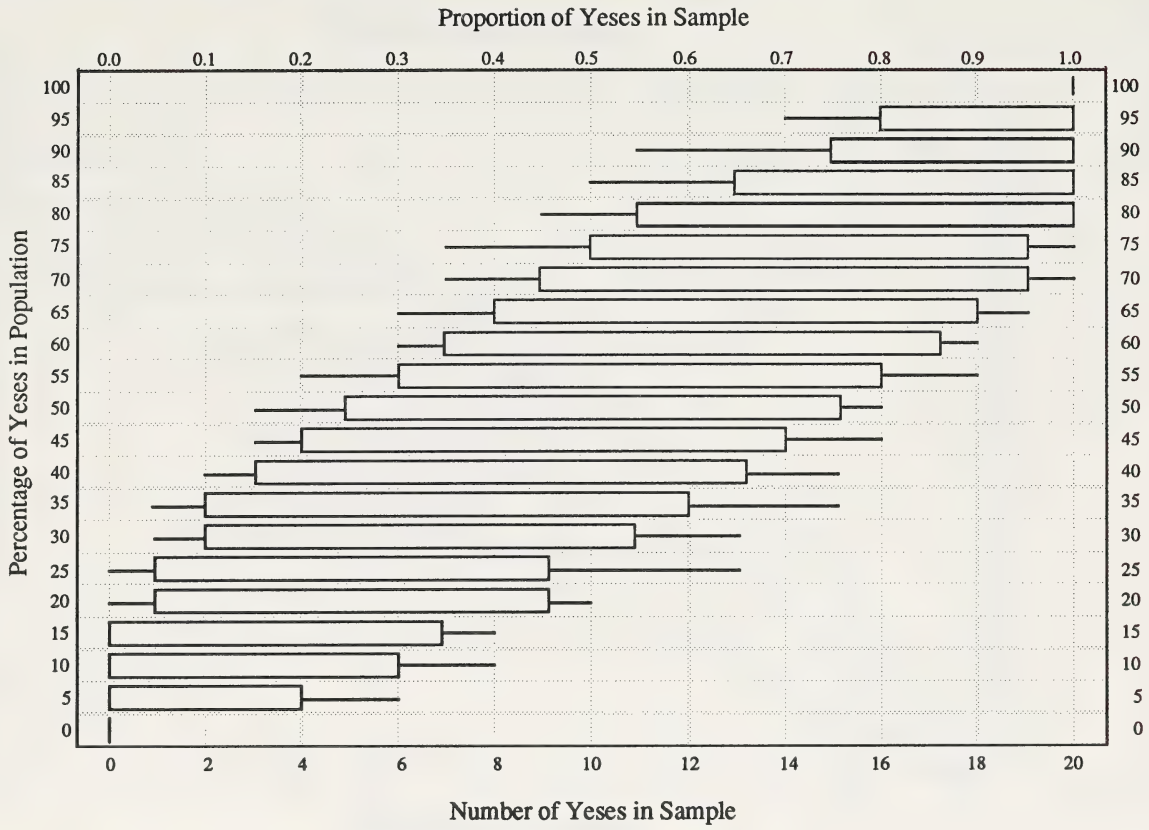


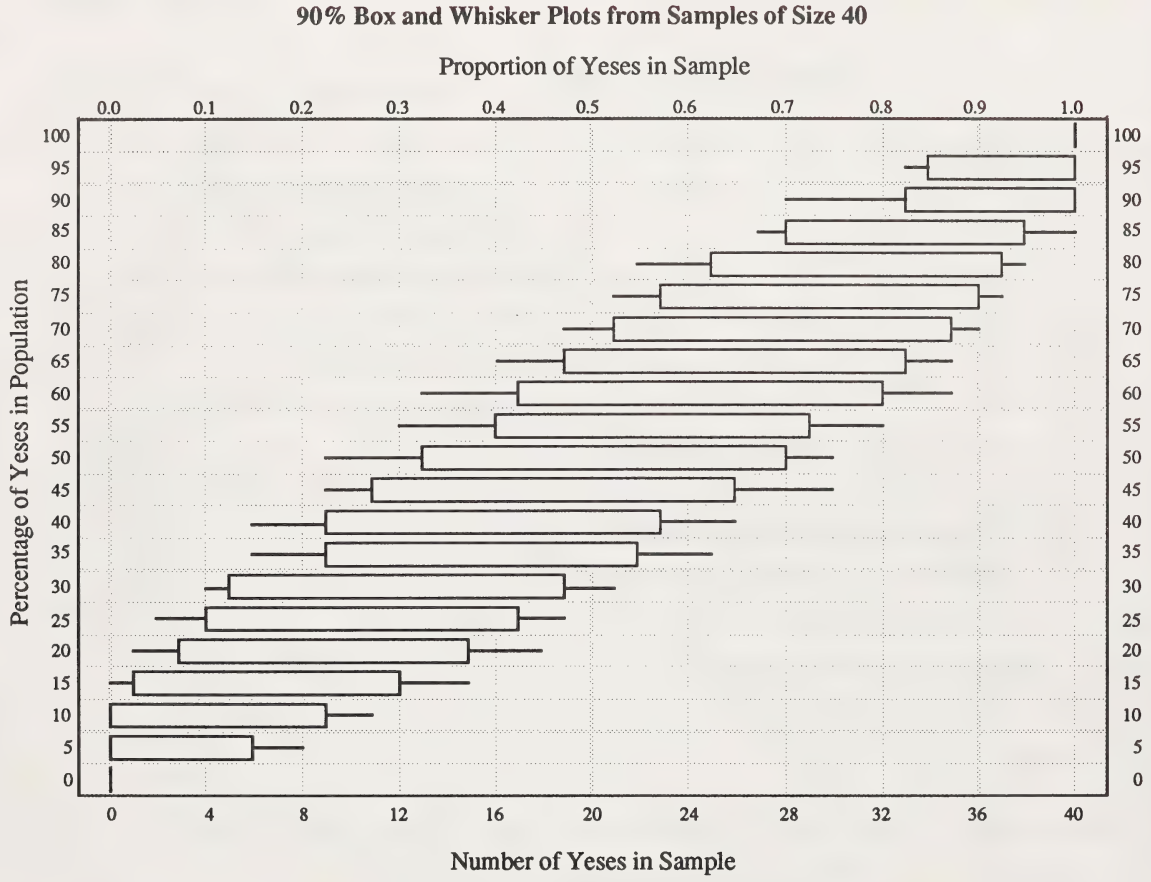
Extensions

1. Answers will vary.

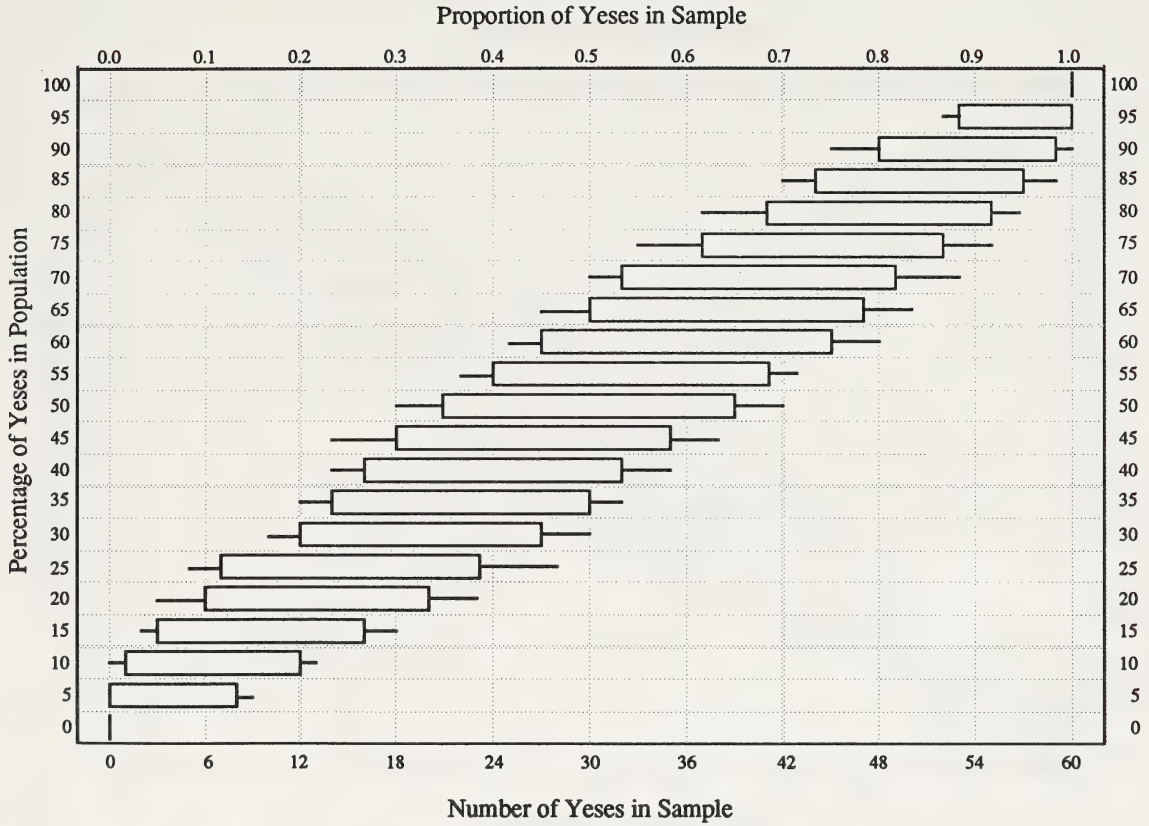
2.
```
100 CLEAR
110 DIM N(60), S(60, 20), S$(60, 20)
120 FOR A = 1 TO 60
130 FOR B = 1 TO 20
140 S(A, B) = INT(100*RND (1))
150 IF S(A, B) < 43 THEN S$(A, B) = "Y"
160 IF S(A, B) >= 43 THEN S$(A, B) = "N"
170 IF S(A, B) < 43 THEN N(A) = N(A) + 1
180 NEXT B
190 NEXT A
195 PR#1
200 FOR A = 1 TO 60
210 PRINT A;"  ";
220 FOR B = 1 TO 20
230 PRINT S$(A, B);
240 NEXT B
250 PRINT "  "; N(A)
260 NEXT A
265 PR#0
```

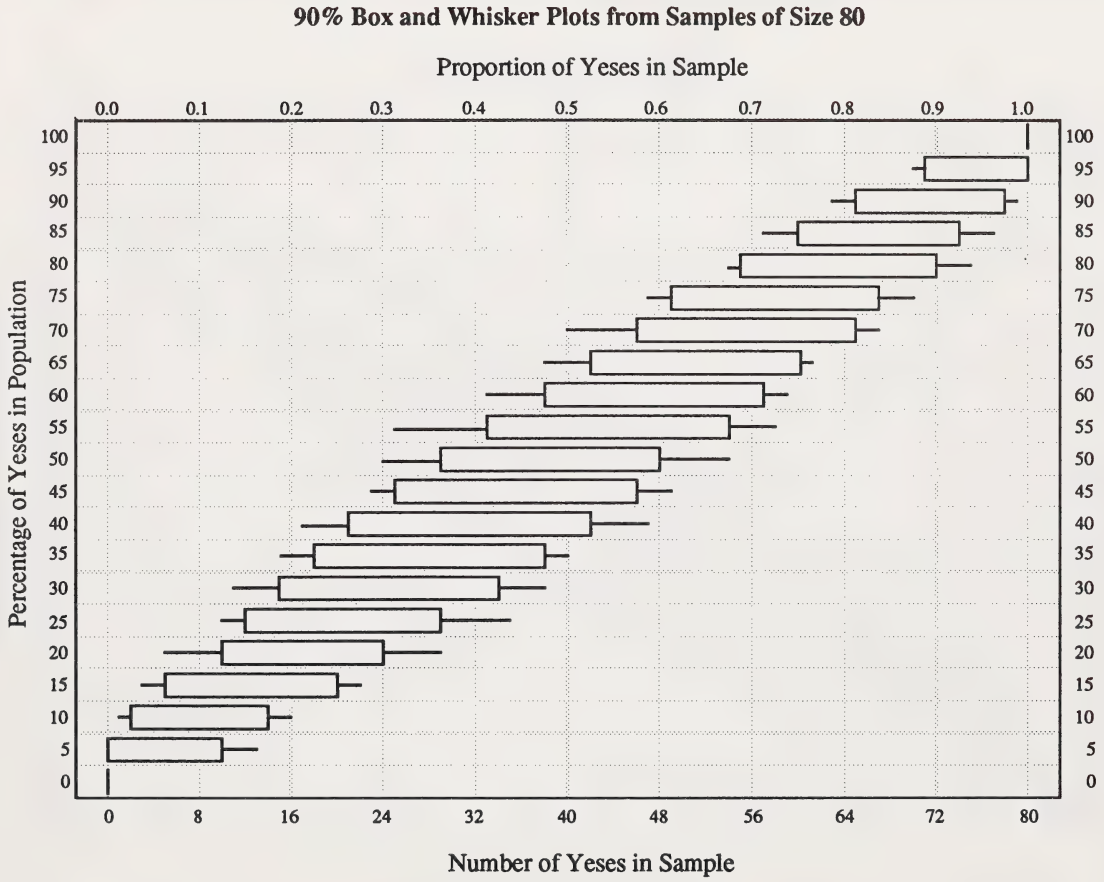
90% Box and Whisker Plots from Samples of Size 20

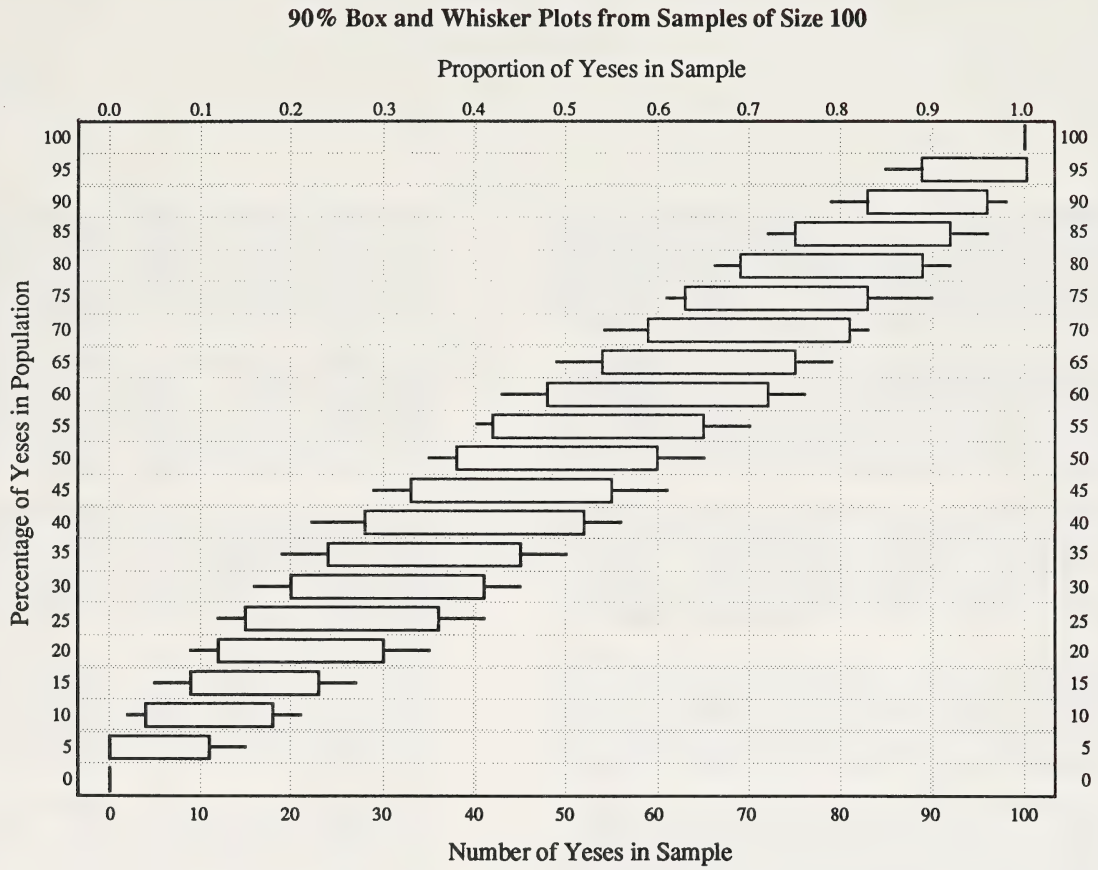



90% Box and Whisker Plots from Samples of Size 40
Proportion of Yeses in Sample
0.0 0.1 0.2 0.3 0.4 0.5 0.6 0.7 0.8 0.9 1.0
Percentage of Yeses in Population
100 95 90 85 80 75 70 65 60 55 50 45 40 35 30 25 20 15 10 5 0
0 4 8 12 16 20 24 28 32 36 40
Number of Yeses in Sample

90% Box and Whisker Plots from Samples of Size 60




90% Box and Whisker Plots from Samples of Size 80
Proportion of Yeses in Sample
0.0 0.1 0.2 0.3 0.4 0.5 0.6 0.7 0.8 0.9 1.0
Percentage of Yeses in Population
100 95 90 85 80 75 70 65 60 55 50 45 40 35 30 25 20 15 10 5 0
0 8 16 24 32 40 48 56 64 72 80
Number of Yeses in Sample


90% Box and Whisker Plots from Samples of Size 100
Proportion of Yeses in Sample
0.0
0.1
0.2
0.3
0.4
0.5
0.6
0.7
0.8
0.9
1.0
Percentage of Yeses in Population
100
95
90
85
80
75
70
65
60
55
50
45
40
35
30
25
20
15
10
5
0
0
10
20
30
40
50
60
70
80
90
100
Number of Yeses in Sample

Random Number Table

91354 55254 27245 38952 64544 74382 78774 82633 63700 11314
89780 09210 30185 98209 95035 20617 39446 64982 36050 98573
27573 59612 00134 87398 42862 93457 84463 01818 22380 11567
85137 59784 93526 87997 11242 40282 00593 46117 23918 66144
56332 03926 00418 11864 18511 44365 23169 36409 91484 10628

93838 41029 52745 26477 72391 52073 04500 41550 41159 28801
50220 01681 93898 96194 96210 44203 49853 02769 78170 18268
92818 03390 12757 30157 35202 67082 74146 13592 07090 09843
95596 25049 70303 52454 96009 64069 57465 03297 31494 18575
36063 66160 71932 35178 97509 73058 02532 14495 71203 42038

50401 31459 63598 59996 36102 80631 22257 90489 50216 46515
24921 06543 54909 98869 93570 48474 10488 13014 33482 06693
26970 15155 97300 29520 40641 52683 74093 56001 15617 89499
03418 73715 70410 26314 14382 01588 29490 45973 89707 97708
99846 83760 54789 92135 72547 22713 91763 44714 49308 13937

78565 27886 94053 21915 52995 34506 09472 94823 01776 10648
38589 59840 74622 13424 91323 43460 23297 93790 96048 16809
65678 04367 76160 65951 74818 32244 51635 87207 93009 56798
65594 07462 70842 32503 35867 29481 31272 96743 31550 02222
58682 02625 74574 16142 12940 76790 66263 23521 68466 02228

38899 83762 78294 26089 32283 33713 63133 92494 95298 82975
25431 00352 12471 34215 88435 86168 55537 28478 60258 31856
60000 15504 22937 97555 26697 49032 46497 79438 48357 07122
28015 92526 61170 37459 55681 78499 21341 27672 97859 38511
06661 00642 17154 91001 43063 34820 09471 54505 56189 09531

68077 55470 47947 75295 72665 61065 98082 49134 03419 36994
90730 34887 80069 30309 65984 68523 25963 77223 96459 88451
57121 35920 88850 30834 96305 44370 10497 49903 12074 03141
40947 63264 82854 83210 48517 22220 85993 58779 91855 07072
11829 65082 46505 47443 02753 81203 56467 67581 17265 37885

10761 56677 19045 50345 44989 64176 91007 76083 67896 20088
74370 97699 12936 80303 32713 30701 14833 96903 41043 56321
11141 13473 99854 79327 68809 85414 74192 10832 35745 25258
57714 76111 17373 72975 61754 47411 78203 91292 29386 01766
68386 86001 74483 03805 75282 51273 45781 13568 00022 39970

15943 19318 59064 58960 95619 95354 10766 89425 72510 84821
06941 32177 75432 07525 69419 77968 63375 92745 91617 02807
39567 37044 65092 88266 90539 87084 31110 66798 50182 47436
57207 92047 88816 13394 38751 34160 27247 28788 25502 00101
75834 68796 73181 78964 98546 47753 20025 12283 02032 51970

88356 89347 49675 97541 11138 67509 00662 13160 83544 17504
74041 47929 39515 93757 37788 04174 53324 29696 22865 95255
66141 58067 62287 28381 37590 56245 55181 12419 22316 16544
55850 90711 71546 41536 87834 98883 59651 64354 82977 47660
35709 94015 45483 30674 00069 85725 40983 57283 05911 36194

10597 10379 85808 03948 09996 29201 03293 54644 73461 17149
29416 31377 02954 74594 44311 49999 82170 56695 63754 19427
15609 56207 90731 86958 92633 11295 91996 08578 76527 38347
99635 44279 18569 89785 37629 78369 97084 36882 73415 64786
81512 77567 94451 97406 21469 69244 09775 05362 60808 86602

N.L.C. - B.N.C.
3 3286 10812143 1

Mathematics 33

9MA33P26

L.R.D.C.
Producer

SECOND EDITION
1991